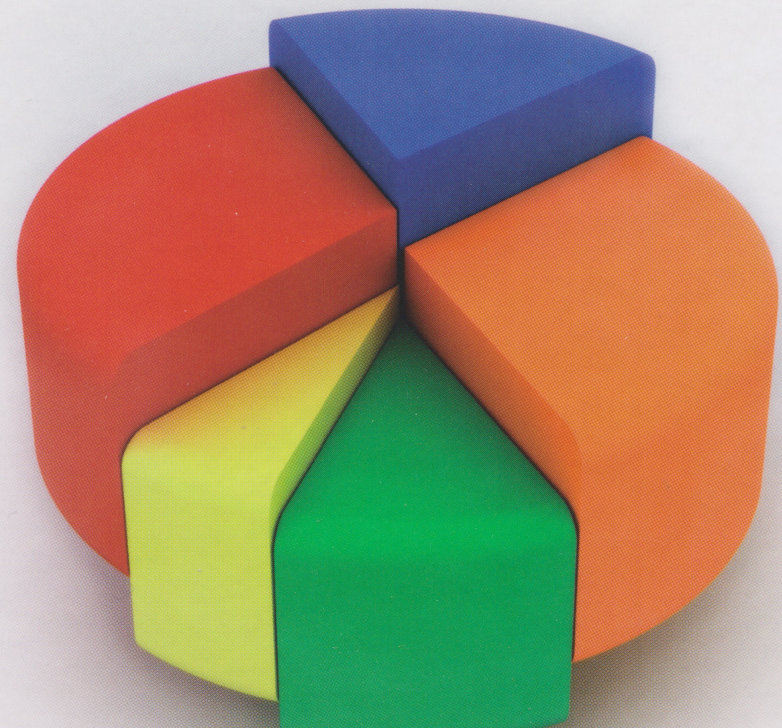
Risks and Hypotheses

The mathematics of economics



Everything is mathematical

Risks and Hypotheses

Risks and Hypotheses

The mathematics of economics

Lluís Artal – Josep Sales

Everything is mathematical

© 2010 Lluís Artal and Josep Sales (text)
© 2013, RBA Contenidos Editoriales y Audiovisuales, S.A.U.
Published by RBA Coleccionables, S.A.
c/o Hothouse Developments Ltd
91 Brick Lane, London, E1 6QL

Localisation: Windmill Books Ltd.
Photograph credits: Corbis

All rights reserved. No part of this publication can be reproduced, sold or transmitted by any means without permission of the publisher.

ISSN: 2050-649X

Printed in Spain

Contents

Preface 7

Chapter 1. The History of Using Numbers in the Economy 9
Numbering systems: The human need to count 9
Algorithms for calculating 17
Basic operations: trade and arithmetic calculation 18
Standard measuring systems 23
Mathematics and economic theories 24
Political arithmetic or the birth of statistics 26
Demographic tables: the second birth of statistics 28

Chapter 2. Money and Inflation 31
A brief history of money from commodity money to fiat money 31
Monetary standards, fiat value and units of account 32
Currencies and exchange rates: why are some currencies worth more than others? 34
The simplest and most practical algorithms 37
The sigma operator 39
How is inflation measured? 44
Interest rates: the cost of money 48
The different interest rates of central banks 49

Chapter 3. Banking and Insurance: Loans and Interest Rates 53
Simple interest and compound interest 53
How long will it take to double the invested capital? 56
How to calculate the instalments on a loan 58
How to reduce the instalments on debts 62
Fundamentals of insurance 66
Descriptive statistics 67

Chapter 4. The Costs of Production and Returns on Investment 69
The optimal production level 69
How and where to save: methods for evaluating investments 78
How are wages determined? 83

The study of unemployment 88
Is it profitable to sign a star player? 92

Chapter 5. How to Buy at the Lowest Possible Price: The Market 97
Supply and demand 97
The theory of market equilibrium 100
Elasticity of demand 105
The operation of the market 107
A market study 111

Chapter 6. Mathematics and the Stock Exchange 117
Stock exchanges and stock markets 117
Fundamental analysis 121
Technical analysis 125
Measuring the performance of the stock market 129
Correlation and regression 132

Chapter 7. Growth and Development 137
Gross domestic product 137
The distribution of production by economic sectors 141
Economic cycles 143
Measuring the well-being of a country 146
Indicators of wealth, economic development and human development 151

Appendix 153

Bibliography 155

Index 157

Preface

The relationship between mathematics and the economy was born of the human need to count, which from the outset was intrinsically related to our survival. There was a need to count the number of members in a family or clan, or the food available to survive. When the clan accumulated more elements than were required, it became possible to exchange these with neighbours for other goods. From the very beginning, the need to count was linked to the requirement for instruments for calculation. At the start, people counted in pairs, before going on to use their fingers, and then later, in order to carry out addition and subtraction, multiplication and division, numbers were invented, followed by algorithms for calculating on paper. All this was spurred on by the development of trade and commercial activities, which were perfected over the course of centuries – and still evolve today.

The first chapter of this book describes the evolution of the signs, numbering systems and calculating algorithms in terms of the contributions made by ancient Egypt and the Romans, Indians, Mayans, Arabs and mediaeval Europeans, until reaching the development of trade and the birth of double-entry accounting during the Renaissance. In addition to this, the birth of the modern state required the standardisation of measuring systems. At the same time, the requirements to collect taxes and plan for the public health system led to the birth of demographic and statistical tables, as well as the first calculating machines, such as Pascal's calculator, invented by the Frenchman Blaise Pascal, a forerunner of modern computers.

The second chapter provides a general overview of the birth of money, which is closely related to the development of trade, growing increasingly more complex with the transition from metal money to paper money, from a monetary standard based on metals to the currency standard. Every country has its own currency that is convertible with respect to the currencies of other countries. Its parity depends on factors such as changes in prices within the country, or in other words, inflation. To calculate inflation (or deflation), governments prepare price indices. To control inflation, they make use of monetary policy, which encompasses instruments such as varying the base interest rate at which banks lend money.

The third chapter is focused on the mathematical formulation and solution of the most common banking transactions, loans and mortgages, refinancing loans and the role played by statistics in economic analysis. The fourth chapter covers the functions of the production of a company, and the return on and assessment of investments.

The fifth analyses the structure and behaviour of the market, both in terms of supply and demand and its role in the formation of prices.

The sixth chapter studies the operation of the stock exchange, in addition to the graphical and mathematical tools used to analyse prices and forecast stocks. Finally, the last chapter discusses the preparation of national accounting, the relationships between macro values, and the interrelationship between sectors of production, economic cycles, the distribution of income and indicators of levels of development.

This book considers the use of mathematics in economics, and how the progressive use of scientific methods has allowed the field of economics to develop. From the innovations of Fra Luca Pacioli, the Renaissance mathematician who invented double-entry accounting, to the modern contributions made by mathematics, the application of mathematics as an instrument of economic analysis has provided highly rigorous economic theories that yield reliable results.

On the pages that follow, we shall cover the main mathematical instruments used in the science of economics, such as the basic concepts of numbering, variables, functions, tendencies and derivatives, curves, equations, sequences, progressions, and correlation and regression that make it possible both to understand what is happening in our own economic environment and the criteria that we must follow in making economic investment decisions, such as depositing savings or accepting the terms of a loan. It explains the main mathematical concepts associated with understanding economic issues such as money, inflation, banking and the financial system. It also addresses financial mathematics used in calculations for paying off loans and the profitability of investments, and analysis of production, supply and demand in markets, the caprices of international trade, how to set prices, the capital market, the stock exchange and economic development.

Chapter 1

The History of Using Numbers in the Economy

We have used numbers since the dawn of humanity. To this day, there are Australian and African tribes that possess only five numbers in their language – one, two, three, four and five. When the solution to an everyday problem gives a result greater than five, the word 'many' is used. Every culture has invented a form of expression (be it figures and letters, words and numbers) required by their way of life and needs, and perhaps these people still don't need to refer to quantities greater than five. They live and calculate in the here and now, in the same way that the Earth's inhabitants have done for tens of thousands of years.

Numbering systems: the human need to count

People need to express quantities using words and figures. Words designate quantities in spoken language, while figures provide a simplified manner of expressing them in written words – and are used to carry out the operations involving them. In the earliest days of humanity, humans fed themselves from hunting and gathering wild fruits, presumably like our ape cousins do today. When, for some reason, their sources of food became exhausted, they had to go elsewhere to find more. This period of our past, in which numbers were hardly required, lasted much longer than the entire subsequent era, during which numbers have been in use.

One day, men and women began to tame animals to secure their food supply. Shortly after, they learnt to cultivate the land as well. This took place around 9000 BC, at which time the Earth had around 8 million inhabitants. From then on, people had a greater need for numbers in their life. Situations arose in which it was necessary to count and write using relatively large quantities and express these using words. Shepherds had to tell their colleagues or neighbours how many sheep they had and if it was roughly the same amount as the previous year. They also had to know the amount of time in a year to plan the breeding of domestic animals and the sowing and harvesting of fields. With the passing of time, they also became required to know the value of the annual taxes that had to be paid to chiefs and priests, who were required to maintain a written record of those who had paid and those who had not. These and many other situations created the requirement for expressing quantities and writing them down.

Special signs, or digits, were created to make it easier to write down quantities. At present, we use the ten familiar symbols: 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9.

How did prehistoric shepherds count their sheep? Once they had gathered their flock together, they grouped them into ten, and then into one hundred (ten groups of ten), and then into thousands (ten groups of one hundred), and so on, until they had counted them all. The base of these groups was always ten, and this is why we say that the base, or basic group, of the numbering system is 10. However, not all shepherds did this in exactly the same manner. Each culture developed its own numbering systems, systems which in spite of always being different, had one element in common: their bases were always divisible by five. Hence, different civilisations used bases 5, 10, 20 and 60.

The number five appeared when primitive humans began to count using their fingers, just like small children do today. Some cultures, such as the Mayans, used their fingers and toes to count, and hence used 20 as their base.

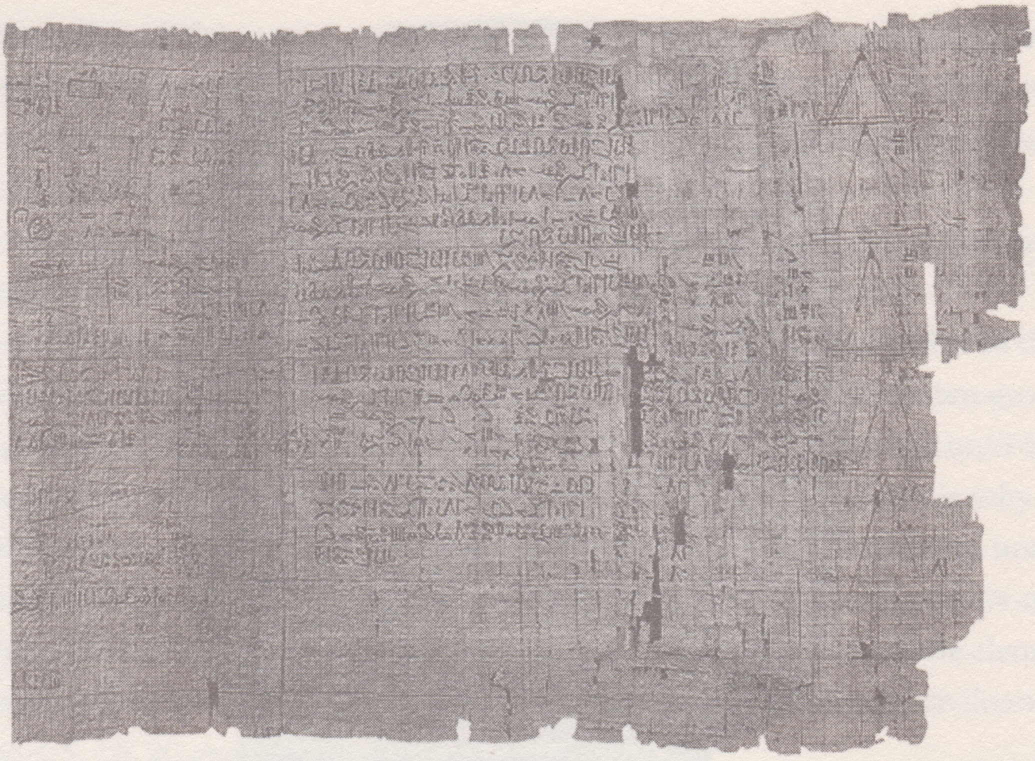
The Sumerians, the Egyptians, the Indians, the Chinese and the Mayans were the first to give numbers a structure and use a numbering system. The Sumerians lived in what is now Iraq around the year 4000 BC. They carried out arithmetic operations and complex geometric calculations to allow them to study the position of the stars in the firmament. These calculations allowed them to create a calendar. The base of the numbering system was 60, and the position of the digits provided numbers with different values: units, tens, groups of 20 and groups of 60. As such, it was is known as a positional system.

Sumerians wrote on clay tablets using reeds with tips cut into a wedge. After writing, the tablets were baked in the oven to harden them. Thousands of tablets containing mathematical calculations have been found under the desert sand.

Sumerian clay tablet from the region of Ur, specifying the areas of a series of farm plots in the city of Umma.

Today we can still see the remains of the Sumerian numbering system when it comes to measuring angles and times. An hour is equivalent to 60 minutes, and a minute to 60 seconds. An angle of one degree (on a calculator: 1 DEG, sexagesimal degree) is composed of 60 subdivisions called minutes (60'), each of which is divided into 60 seconds (60").

The ancient Egyptians wrote on papyrus, made from the pith of a plant that grew on the banks of the River Nile. The priests of this culture made numerous scholarly discoveries about numbers. They invented fractions, primarily used to calculate the third part of the harvested corn to be paid as a tax, or to refer to the parts into which a loaf of bread could be divided. In 1858 the Scottish Egyptologist, Alexander Henry Rhind, bought an ancient papyrus scroll that contained various mathematical problems on fractions and geometry. In fact, this is the first book of mathematics that is known. It was written around 1700 BC and allows us to understand the development of the science in ancient Egypt.

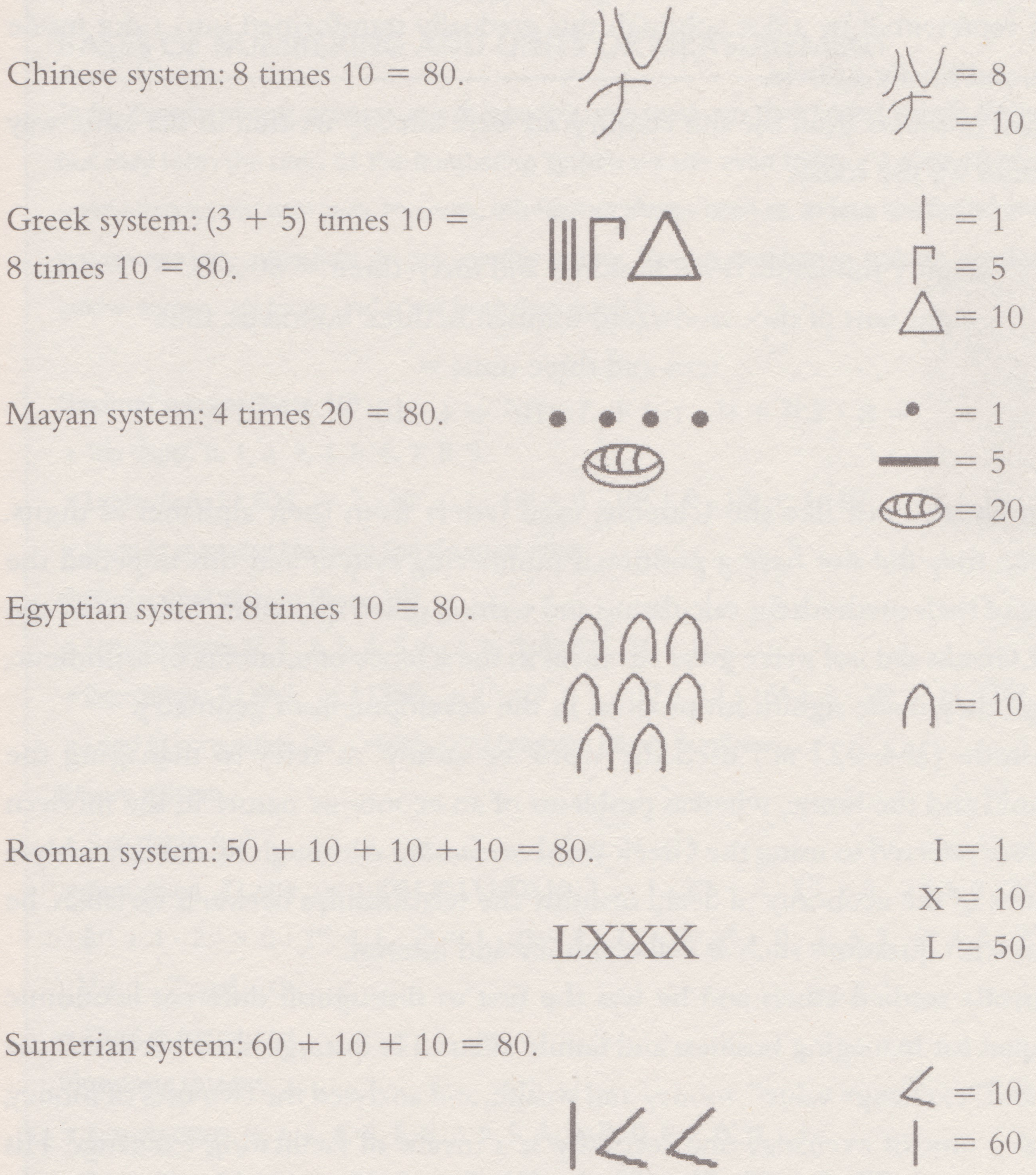


Fragment of the so-called Rhind Papyrus, *the mathematical bible of the Egyptians. The scrolled document is 33 cm high and 5 m long.*

The Chinese, for their part, chose a vertical system of writing. They classified numbers into even and odd, and their achievements include beginning to use negative and positive numbers. As they made use of letters and signs from their alphabet as digits, each word also had a numeric value, which was a source of problems for later analysts. Furthermore, they believed that words had a magical significance that depended on the number they represented, creating a whole new layer of mysticism over the field of mathematics.

The Mayans, who lived in Central America many years before the arrival of Christopher Columbus, also used a vertical system of writing. They developed a calendar with 20-day months and a 360-day year. They used a base 20 positional numbering system and a series of signs for expressing quantities that was very similar to the ones used in China and India.

The Mayans and their Olmec predecessors made great advances in the fields of mathematics and astronomy, defining the concept of zero or 'nothing' around 36 BC C, the first documented reference to that very special number. However, the problem with the Mayan zero was that it could not be used for arithmetic operations and this prevented the Mesoamerican civilisation as a whole from developing advanced calculation methods.


Chinese system: 8 times 10 = 80.
= 8
= 10
Greek system: (3 + 5) times 10 = 8 times 10 = 80.
= 1
= 5
= 10
Mayan system: 4 times 20 = 80.
= 1
= 5
= 20
Egyptian system: 8 times 10 = 80.
= 10
Roman system: 50 + 10 + 10 + 10 = 80.
LXXX
I = 1
X = 10
L = 50
Sumerian system: 60 + 10 + 10 = 80.
= 10
= 60

The same number in six different numbering systems.

The Indians were the most prodigious calculators of the ancient world. The used enormous quantities in their arithmetical calculations and set problems that showed great imagination. (In one, for example, 1,024 female monkeys take part in a battle.)

Two of the greatest mathematical inventions have been dated to the 6th century AD. The first is assigning a different value to each digit depending on its position when writing its quantity (units, tens, hundreds, thousands…). The second is devising a special sign to represent zero for indicating the number of elements in an empty set. The number was called *sunya* in Sanskrit. The Arabs called it *cephir*. Zero was initially represented by a dot although this gradually transformed into a dot inside a circle, and finally a circle.

Indian numbers from the 6th century AD were already written in the same way as the ones we use today.

$$\text{Eighty thousand, three hundred and forty-three} = 80{,}343 = \text{eight tens of thousands, zero thousands, three hundreds, four tens and three units} = = 8 \cdot 10^4 + 0 \cdot 10^3 + 3 \cdot 10^2 + 4 \cdot 10^1 + 3 \cdot 10^0.$$

The Greeks, just like the Chinese, used letters from their alphabet as digits. However, they did not have a positional numbering system and this impeded the progress of their methods for calculating and writing down quantities. This is why the ancient Greeks did not make great advances in the science of numbers or arithmetic, although they made significant progress in the development of geometry.

Aristotle (384–322 BC) used the word 'economy' to refer to managing the household and the home, whereas problems of an economic nature in the modern sense were referred to using the Greek word *crematistica*. Although he did not analyse all matters of the economy in detail or study the relationships between variables, he did deal with questions such as value, money and interest.

Aristotle studied ethics and he was the first to distinguish different economic techniques for managing business and family affairs. He distinguished between "use value" and "exchange value", money and wealth, and analysed the two uses of money, firstly as a unit of exchange and secondly as a means of facilitating exchange. His distinction between interest and usury (and his ethical condemnation of the latter) lasted to modern times and formed the basis of the doctrine of the Catholic Church. The philosopher also reasoned on other economic matters, such as private property and slavery, and his ideas were a considerable influence on Islamic ethics.

The Romans did not make much improvement to Greek numbers. They used the letters M, D, C, L, X, V and I as numerical symbols, writing bars above them to designate larger numbers. This meant that they suffered the same problems as the Greeks. It is easy to see the difficulty of writing a quantity as large as 1 million or carrying out operations using various quantities in the Roman system.

This meant that when the Arabs brought the Indian system to Europe via Andalusia in the 8th century AD, those working on calculations began to use it immediately and abandoned the Roman numbering system for good.

BASES OF NUMBERING AND UNITS OF MEASUREMENT

Today it seems almost inconceivable that people once used only their fingers to help them count, but they form the basis of the numbering system we use even today – a base 10 positional numbering system with zero. However, this has not always been so. In fact, the fastest and most precise modern computers do not use this system. Various numbering systems are described below, some used today and some from days gone by.

Decimal system (base 10)

- Ten digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
- Expression: $72{,}603_{10} = 7 \cdot 10^4 + 2 \cdot 10^3 + 6 \cdot 10^2 + 0 \cdot 10^1 + 3 \cdot 10^0$.
- Used in everyday life since the dawn of time.

Hexadecimal system (base 16)

- Sixteen digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F.
- Expression: $72{,}603_{10} = 11B9B_{16} = 1 \cdot 16^4 + 1 \cdot 16^3 + 11 \cdot 16^2 + 9 \cdot 16^1 + 11 \cdot 16^0$.
- Used in computing (as a means of shortening binary numbers).

Binary system

- Two digits: 0, 1.
- Expression: $72{,}603_{10} = 10001101110011011_2 = 1 \cdot 2^{16} + 0 \cdot 2^{15} + 0 \cdot 2^{14} + 0 \cdot 2^{13} + 1 \cdot 2^{12} + 1 \cdot 2^{11} + 0 \cdot 2^{10} + 1 \cdot 2^9 + 1 \cdot 2^8 + 1 \cdot 2^7 + 0 \cdot 2^6 + 0 \cdot 2^5 + 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$.
- Used in computing.

Vigesimal system

- Twenty digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, G, H, I, J.
- Expression: $72{,}603_{10} = 91A3_{20} = 9 \cdot 20^3 + 1 \cdot 20^2 + 10 \cdot 20^1 + 3 \cdot 20^0$.
- Used by the Mayans and Sumerians with their own symbols.

It should also be noted that throughout history different cultures have made use of a range of systems of units of measurement for magnitudes (weight, length, money, volume), which are related to their numbering systems. In general, the writing and calculations used in economics can be considerably simplified by using equal divisions and partitions to designate multiples and submultiples for units of measurement and in expressing quantities. The decimal metric system is a suitable choice for this. It defines multiples of 10 to designate the multiples of a unit of measurement and also employs the decimal numbering system to express quantities (eg. 2.547 metres is 2 metres, 54 centimetres and 7 millimetres).

ANCIENT AND MODERN MULTIPLICATION

The procedures, techniques and methods for calculating and obtaining the results of arithmetic operations are referred to as 'algorithms' and there have been enormous advances in this field throughout the history of humanity. The following tables represent two algorithms for carrying out the multiplication 2,409 · 94, according to the methods in use at two periods in history.

2,409 × 94 using the method used by the ancient Egyptians (3600 BC)

2,409	94
(1,204	188)
(602	376)
301	752
(150	1,504)
75	3,008
37	6,016
(18	12,032)
9	24,064
(4	48,128)
(2	96,256)
1	192,512

= 94 + 752 + 3,008 + 6,016 + 24,064 + 192,512 = 226,446.

The first factor (2,409) is successively divided by two until reaching one. In parallel to this, the second factor (94) is doubled as many times as the first factor is halved. The result of the product is given by adding all the numbers in the right column with an odd number to their left.

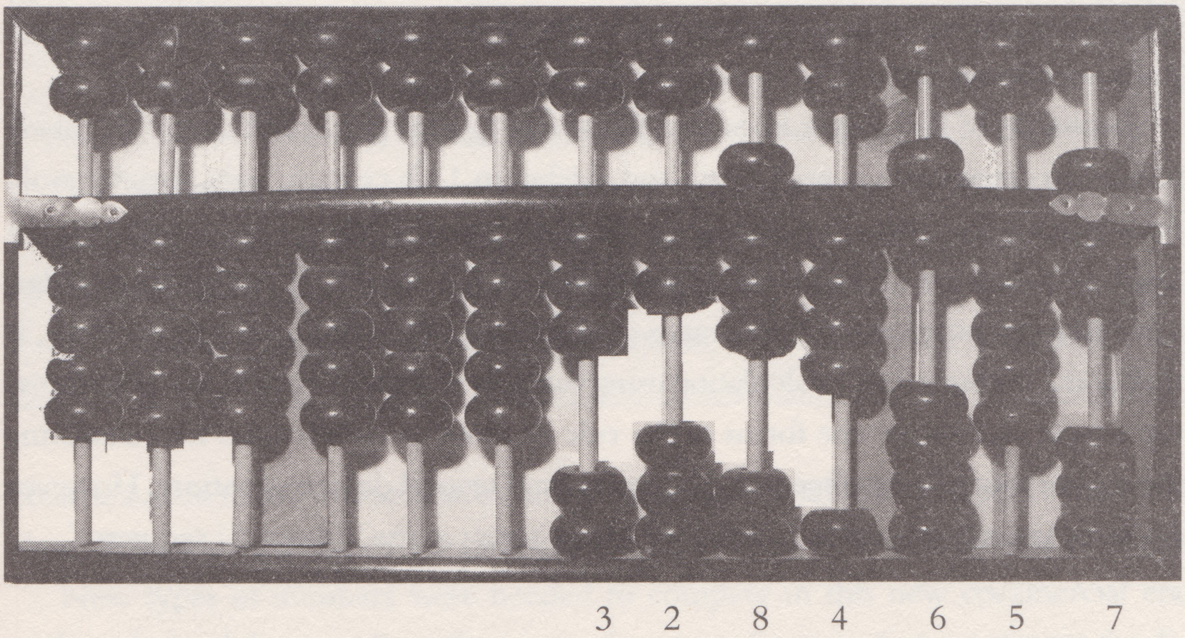
2,409 × 94 by the method used by a computer in the 21st century

					9 4	
					x	
		2	4	0	9	
				3	6	4 · 9
			8	1	0	90 · 9
					0	4 · 0
					0	90 · 0
		1	6	0	0	4 · 400
	3	6	0	0	0	90 · 400
		8	0	0	0	4 · 2,000
1	8	0	0	0	0	90 · 2,000
2	2	6	4	4	6	

Multiply 4 by 9, 90 by 9, 4 by 0, 90 by 0 and so on… until 90 by 2,000. The result is the sum of the partial results listed under the operands.

Algorithms for calculating

The Greeks and Romans calculated in the same way as prehistoric shepherds, using piles of pebbles or small sticks. Numbers were only used to write down the results. The word *calculate* comes from the Latin word *calculus*, which means 'small stone'. In order to avoid having to always carry stones in their pockets, they invented the abacus, a device still used by Asian merchants to calculate prices.



A modern abacus with the number that is represented on it.

Each column on the abacus is one position of the quantity that is being expressed. If no bead has been moved on the row, this means there is a zero in that position, despite the fact that the Romans could not represent zero in their system of numbers. In the Roman world, the number three million, two hundred and eighty-four thousand, six hundred and fifty-seven is written as:

$$\overline{\text{MMMCCLXXXIV}}\text{DCLVII}.$$

However, in the 5th century AD, the Indians were already writing the same number as 3,284,657 just as it is written now. In the 8th century, the Arabs incorporated the Indian inventions of positional numbering and the zero into their culture after their conquest of northern India. During the Middle Ages, the Arabs developed negative numbers, the rule of three and methods of proportionality for solving problems such as: "Hussein has 22 dinars, Omar 19 and Khalil 7. They pool their money together to carry out a business transaction that earns them 12 dinars. How should this profit be distributed?"The Koran also describes complicated inheritances that inspired Arabic law and obliged calculators to develop mathematical procedures for proportional distribution depending on the degree of relation to the deceased.

Algebra was developed to solve these types of problems and equations. The word comes from the Arabic term *al-jabr* which means to 'recompose' or 'reconstruct'. The same period of Islamic scholarship saw the definition of the first algorithms (or calculating procedures), a word rooted in the name of Al-Khwarizmi, the greatest Arabic mathematician.

Leonardo of Pisa, better known as Fibonacci (meaning 'son of Bonifacio'), made numerous discoveries in arithmetic and algebra in 13th century Italy, which would subsequently be developed during the Renaissance (14th and 15th centuries). His *Book of Calculation* (*Liber Abaci*) brings together the wealth of Arabic knowledge, including the positional numbering system and zero (which he called *zephyrum*), as well as the techniques for calculating using integers and fractions. The text also shows the simple and composite forms of the rule of three, as well as rules for calculating the square root of a number and solving first and second degree equations. However, his most famous invention is what we know today as the 'Fibonacci sequence'.

Basic operations: trade and arithmetic calculation

The first book on commercial arithmetic was published in Treviso (Italy) in 1470 and was written by an anonymous author. Throughout the course of the 15th century, around 30 books were published on the subject (14 in Italy and 11 in Germany alone). All contained a description of the Arabian base 10 numbering system and the procedures or algorithms for the four operations on integers and positive numbers (so-called natural numbers). They also included the description of fractions and their operations, the rule of three, progressions, how to solve problems directly applicable to commercial situations (such as calculating the real value of a good being exchanged), the calculation of taxes and tolls, the use of alloys and the conversion of units of measurement.

The period also saw the formation of a closer relationship between the economy, which was understood as the management of scarce resources, and mathematics, defined as abstract thought based on the correctness of logical deductions using basic arithmetic. Being able to add and subtract numbers (in reference to goods that are exchanged) became a basic practice. Subsequently, the development of trade

CLASH OF ALGORITHMS

In Christian Europe, the advances of the Middle Ages were limited to translating the original Arabic works and the Arabic versions of ancient Greek books, such as Aristotle's *Economy*. Shortly before the year 1000, the monk Gerbert d'Aurillac, who would go on to become Pope Sylvester II, learnt Arabic numerals and their positional value from Andalusian Arab scholars and made improvements to the Roman abacus, although it continued to lack a zero. It was not until the 12th century that the Crusades would import the Indo-Arabic zero and calculation methods from what was then the Kingdom of Jerusalem. However, the church attempted to prevent the use of Arabic calculating methods, arguing that if they were simpler and more effective, it was because they were connected to magic or worse – the devil. At that time, professional calculators (mathematically minded people) used these algorithms in secret to avoid problems with the authorities. However, in spite of ecclesiastic oppression, Arabic numbers were strongly advocated in the pragmatic and Earthly mercantile realm from the start of the Renaissance.

created a need to use multiplication and division. As a result efficient procedures and algorithms were developed for carrying out these operations quickly and precisely.

New types of numbers were needed to compete in this new commercial era, such as decimal fractions for referring to the parts of a unit of measurement for dimensions, weights and volumes, as well as for expressing their values using currency or barter. However, such numbers had already appeared in Egyptian hieroglyphs.

An extremely important application of centesimal fractions was representing the percentage as a description of the parts of a whole when it came to discounts and interest. Later, this new type of number would be generalised to include other fractions with denominators other than 10 or 100. The historical evolution came to an end when the quantities were expressed using the positional base 10 numbering system, which we still use for everyday applications.

$$36/100 \text{ (centesimal fraction)} = 3.6/10 \text{ (decimal fraction)} =$$
$$= 36\% \text{ (percentage)} = 0.36 \text{ (decimal number)}.$$

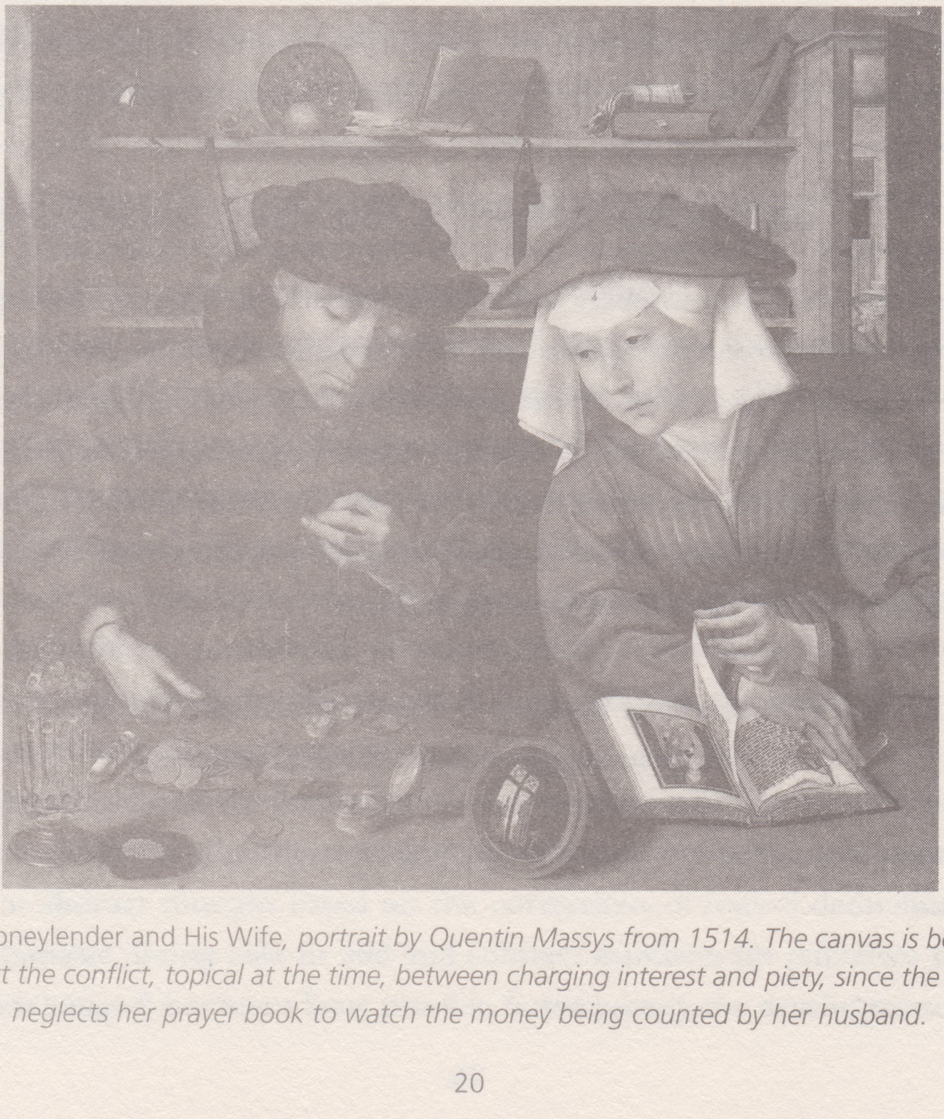
In the Renaissance, the quantity 78, 4/10, 5/100, 6/1,000 was written as 78 + 4/10 + 5/100 + 6/1,000 = 78.456.

Fractions of this type were used from the 17th century and were referred to by mathematicians as rational numbers. They can be written in two ways (or in two notations), as a fractional-percentage or as a positional decimal, using a point.

Rational numbers can have a finite (limited) number of decimal places. This is the case with a precise division, such as 34/64 = 0.53125.

And they can also have an infinite (unlimited) number of digits with a known repetition, or period, such as 34/70 = 0.4857142857142857142857142857…

During this era, banks began to appear to ensure the circulation of money that accompanied the transactions of goods and services took place under secure conditions. The first bankers were mediaeval metalworkers who were frequently Jewish or Muslim. Christians were prohibited from lending money because the church regarded it as the sin of usury – and punished any sinners. However, history would record that many Christians subsequently became bankers and moneylenders.



The Moneylender and His Wife, *portrait by Quentin Massys from 1514. The canvas is believed to reflect the conflict, topical at the time, between charging interest and piety, since the woman neglects her prayer book to watch the money being counted by her husband.*

THE INVENTION OF ACCOUNTING

In 1494 the Franciscan friar Luca Pacioli invented modern accounting. His great idea was the concept of double entry, a magnificent example of the contribution mathematics has made to trade. From that point onwards, the need to reflect income and expenditure was no longer meaningless but came to be seen as a process of cause and effect. His book *La Summa di Arithmetica, Geometria, Proportioni et Proportionalità* was published in Venice in the context of the new political and commercial relationships that were coming into force at the time. Written as a popular textbook, it describes the operation of a mercantile business prior to the initiation of the operations of buying and selling goods. It advises businessmen to make a complete list of all their property (assets) and another of their commitments to third parties (liabilities). Assets should be grouped according to the principle of value and mobility (cash first, then assets). Liabilities should be grouped according to their expiry and payment term (short-term and long-term). Keeping records using the double-entry system shows that all mercantile operations that take place have a cause, which must necessarily produce an effect, meaning that there is numerical compensation between the cause and the respective effect.

On the accounting concept of goods, Pacioli wrote: "For everything you include in it, you will record yourself as debtor day by day in your books and hence also conversely, make yourself the common creditor of all that you draw or obtain from it, and as if it were a debtor that pays you in part." The warehouse of goods is regarded as a debtor of everything placed inside it or spent on it for any reason. Pacioli proposed four basic books to record transactions: inventory (book of inventories and balances), *giornale* (general journal), ledger (general ledger) and *memoriale* (which no longer exists). The text also explained the importance of using accounts and other accounting concepts to record the performance of the business and allow the merchant to decide if it was functioning correctly.

Accounting rose out of a growth of trade that occurred during the early Middle Ages and the Renaissance (15th and 16th centuries) to allow businessmen to state volumes of trade and profits clearly and calculate the value of invested capital and assets. Embarking upon a new business venture that required extra capital meant that an entrepreneur would be required to take out a loan. From that point onwards, banks and double-entry bookkeeping became essential mathematical tools.

This saw the creation of negative numbers, also known appealingly as 'debt numbers' or 'absurd numbers'. They received this second name because they arose at a time when people were struggling to cope with the here and now. Something as esoteric as a number less that zero was truly absurd in this context, almost meaningless (but not to creditors, of course). 'Absurd numbers' were revealed in India in the 5th century AD AD but did not reach western thought until the 16th.

Negative numbers were also handled using the abacus, but the bars or counters had a different colour to those used for positive numbers. By the 5th century the use of the Germanic symbols + and − to indicate the surplus or lack of merchandise in warehouses was becoming common practice among merchants. In the 16th century they were still referred to as 'false' numbers and negative numbers were not universally accepted until the end of the 18th century.

The expansion of the numbering system to include negative numbers was extremely practical for simplifying the class of operations for accounting. With the incorporation of negative numbers and the use of rational numbers, it became possible to find a mathematical solution to any first-degree equation.

For example, the equation $P + 50 = 32$ did not have a solution until negative numbers were accepted: $P + 50 - 50 = 32 - 50$; $P = 32 - 50$; $P = -18$.

With the rise of the Protestant church in the 16th century, Lutheran believers and Calvinists broke free from the prohibition of usury. This created the possibility of generalising these mercantile practices and facilitating the accumulation of capital, with the appearance of the so-called 'money ethic'. The era saw the birth of capitalism. This was when the idea of accruing wealth became an important goal; obtaining money ceased being a means and instead became end in itself. There was

The sociologist Max Weber (centre, in 1917) wrote The Protestant Ethic and the Spirit of Capitalism *in 1905, a fundamental text that sets out various hypotheses on the relationships between Lutheran morality and the development of capitalism.*

"DEBT NUMBERS"

One of the first definitions of a negative number was provided by Father Tomás Vicente Tosca in his work *Compendio Matemático*, published in 1709. One of the problems set out in the text describes a common situation in the use of capital using relative quantities, referred to as having and owing, or debt and profit, in order to give negative quantities a meaning and illustrate how a quantity that is less than nothing is to be understood. In one of his chapters he explains: "Imagine someone who does not have any assets and owes 1,000 escudos, and someone else who has no assets but owes nothing. It is clear that the former is less fortunate than the latter, despite both having nothing; hence the former must have less than nothing. Similarly, if we give 1,000 escudos to the person who has nothing and owes 1,000 escudos, allowing him to pay off his debt, he will increase his assets. However after this increase his assets are nothing; hence prior to this increase, his assets must have been less than nothing."

a paradigm shift: to behave rationally or normally was to work in such as way as to obtain the maximum quantity of wealth.

The new Protestant ethic had a very characteristic distinguished feature - an asceticism that explains why many parsonages and church buildings became commercial hubs. Thanks to their capitalist head for business, Protestants assumed an intense form of devotion that impregnated and regulated every aspect of their lives. For the Protestant ethic, enrichment was a means of ensuring eternal salvation.

Standard measuring systems

In the 18th century, the era of enlightenment and revolutions, the development of commercial activity urgently required the use of instruments for physical measurement, as well as the development of more advanced currency systems. In 1791, the French national assembly defined the metre as a ten-millionth of the quadrant of the Earth's meridian and the kilogram as based on the mass of a litre (0.001 cubic metres) of water. Thus was established the first system for weights and measures to facilitate international exchange – the decimal metric system.

When the calculations and measurements of the Earth's meridian became more accurate thanks to our calculations of the size of the Earth, and the properties of water were studied in greater depth, certain discrepancies were found in the previously defined standards. In 1799 the metric system was re-established based on standards for the metre and the kilogram manufactured from a platinum alloy.

This marked the beginning of the Industrial Revolution. The standardisation of mechanical parts, particularly screws and nuts, was a matter of great importance and required measurements of increasing precision. These requirements saw the relationship between standards and their natural inspirations begin to dissolve. From that point onward, the standards themselves became the basis of the measuring system, and this remained the case until the 1960s.

One the great advantages of the metric system is that multiples and submultiples are decimal multiples and divisors. Prior to this, the units used were divided into three, twelve or sixteen parts depending on the unit and the country of origin, frustrating arithmetical operations using algorithms based on the decimal numbering system.

The metric system was internationally adopted at the 1889 General Conference on Weights and Measures, which became the International System of Units as defined in 1983. At present, approximately 95% of the world's population lives in countries that use the metric system and its derivatives.

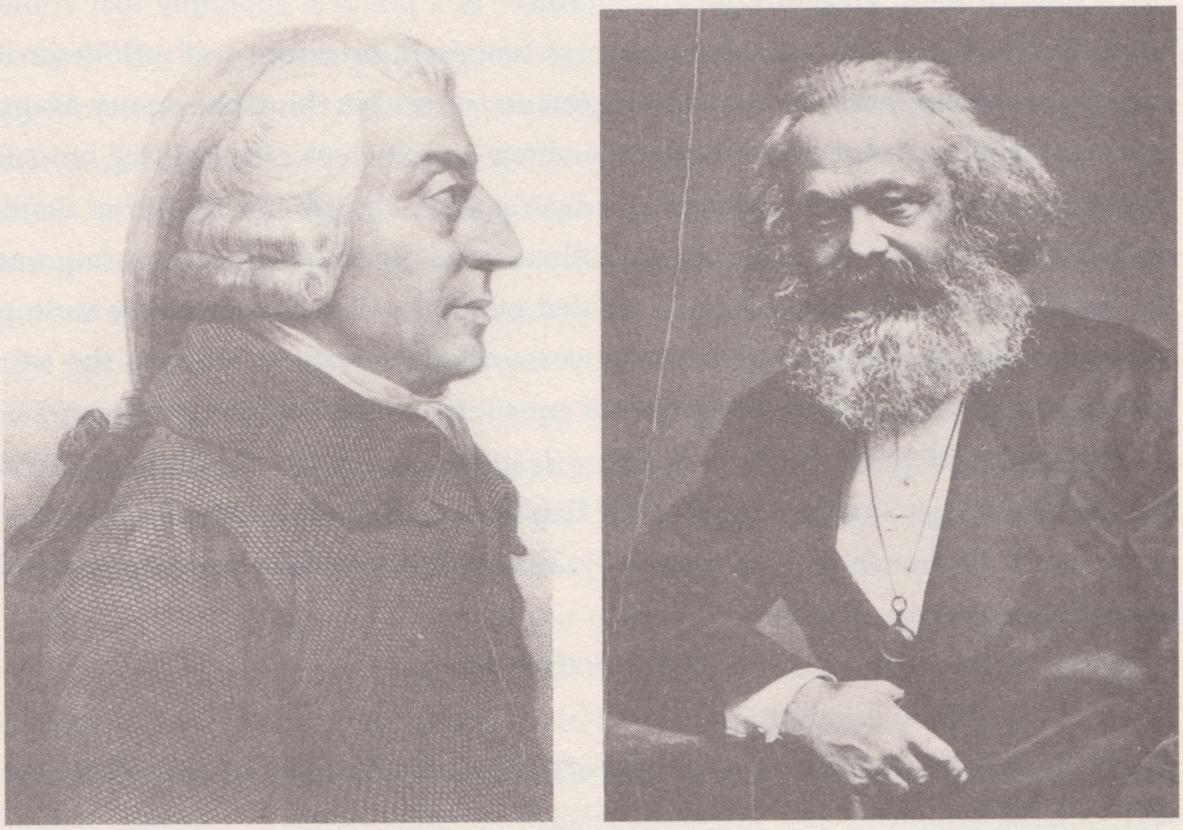
Mathematics and economic theories

Different schools of economic thought arose between the 16th and 18th century, the objective of which was to facilitate the enrichment of the state, hence the reason why the science of economics is often referred to as 'political economy'. This saw the birth of mercantilism, which holds that the state can only become wealthy if it sells more to other states than it purchases. Hence it must have a positive balance of trade that allows an net influx of precious metals. The major contribution of this ideology was the formulation of the quantitative theory of money.

In the 18th century the Physiocrat school analysed society in a similar way to the human body, through the prism of 'natural order'. Their analytical spirit led them to provide a precise definition of the structure of the economic system of a country, as can be found, for example, in the French medic François Quesnay's *Tableau Économique*, dealing with industrial relations, a precursor to the input-output tables of the future.

In 1776, Adam Smith, regarded as the father of political economy, published his *Wealth of Nations*. The innovative theoretical aspects of the work were the distinctions between use value and exchange value; the recognition of the division of labour, understood as specialising in specific tasks to reduce production costs; the prediction of potential conflicts between factory owners and their poorly paid workers; the accumulation of capital as the source of economic development; and the defence of the market as the most efficient mechanism for allocating resources. As such, Smith was a proponent of economic liberty.

Karl Marx, on the other hand, claimed that only labour produced value, an argument that is developed at length in his book *Das Kapital* (1864). Marx establishes four different concepts of value: individual (for comparing direct value and production value), direct (only accounting for competition within the sector), production (accounting for competition within and between sectors) and effective (the real market price). However, Marx never managed to formulate a mathematical analysis of the problem.



Adam Smith (left) and Karl Marx, two thinkers who made great contributions to political economy from completely different viewpoints.

Throughout the 19th century, many economists such as Ricardo and Marx, developed the theory of prices and value. However it was only thanks to the work of subsequent thinkers, such as Walras, Sraffa, Menger, Von Neumann and Morgersten, that the ideas became mathematically defined and mathematical economy arose. Léon Walras criticised the liberal theories in vogue at the time as being inadequate when it came to explaining the economic problems of the time. In *Elements of Pure Economics* (1874) he made clear his disagreement with David Ricardo's theory of the value of labour and the profit from property (land, buildings, etc.) and placed the classical inheritance of Adam Smith in doubt. Together with the mathematician Antoine Cournot, he introduced the mathematical calculus of economics.

The Walras model represented the price of the level of supply and demand in interrelated equations and involved three variables: prices and the levels of supply and demand, meaning two unknowns had to be calculated (price and quantity), since under equilibrium conditions, supply equals demand. Walras was the first to provide a mathematical formulation of general equilibrium, one of the processes related to the analysis of supply and demand.

In spite of these great advances, the project of a political economy that could be written in the mathematical language of functions, equations and infinitesimal calculus was strongly criticised by other great economists in the 19th century. Many made use of historical research methods and rejected the idea of translating human liberty into mathematical numbers and operators as absurd. On the other hand, mathematicians were also reluctant to embrace this type of analysis, pointing out the paucity of results that had been yielded by mathematical attempts at taming economics up to that point. The many years of collaboration between the two disciplines had only yielded the system of equations for the Walras equilibrium.

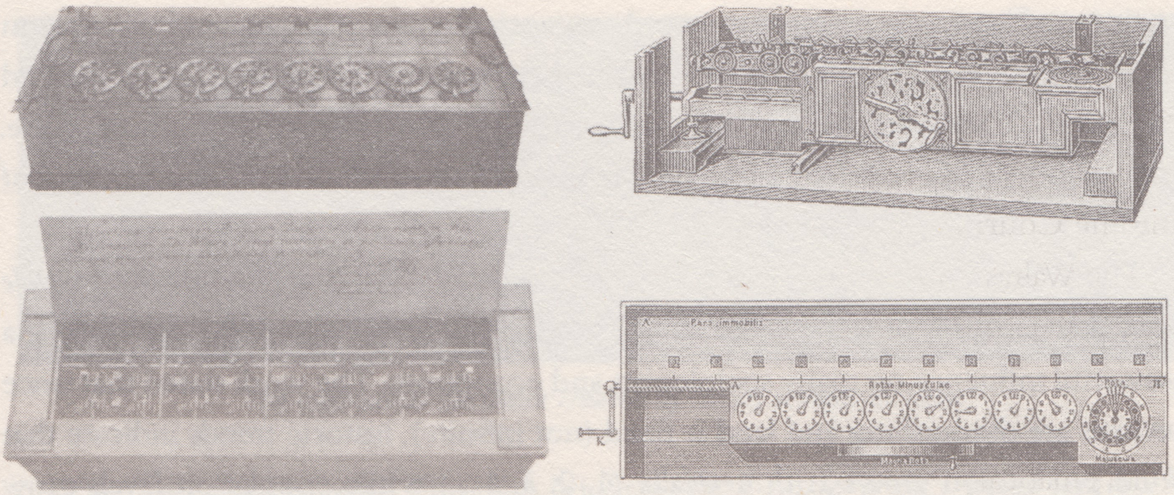
The study of price formation saw a leap forward with the mathematical formulation of Piero Sraffa, a pupil of the English economist John Maynard Keynes. In his book *Production of Commodities by Means of Commodities* he proposed the following equation:

$$P = S + B + R,$$

in which price (P) and the variables of salary (S), profit (B) and rent (R) refer to the state of equilibrium.

Political arithmetic or the birth of statistics

In 1642, a young French mathematician, Blaise Pascal, invented the first calculating machine based on a toothed-wheel mechanism. It could handle numbers of up to eight digits and implemented the operations of addition and subtraction. Pascal created the device to help his father, a tax inspector whose workload had increased, who was. Some fifty machines were manufactured.



Pascal's 'Pascaline' calculator (left) and the 'Stepped Reckoner' device invented by Leibniz (right).

In 1694, the German philosopher and mathematician Gottfried Leibniz made improvements to Pascal's calculator, creating the first machine for multiplication and division. Jakob Bernoulli, from Switzerland, wrote *Ars Conjectandi* in 1705, the year before his death, in which he developed the first theory of probability. In this work he stated something as general as the fact that uncertainty reduces when the number of observations increases. Bernoulli proposed the following ideal experiment: "A large urn contains 3,000 black balls and 2,000 white balls. If balls are taken out and their colour noted down before returning them to the urn, we can see that the more balls that are taken out, the closer the proportion of white and black balls grows to 2:3." Today, the statement is known as the 'law of large numbers', and forms one of the foundations of mathematical statistics.

Antoine Lavoisier, who invented the naming convention and formulation of modern chemistry, was a supporter of the French Revolutionary forces. He was involved in the creation of the metric system and offered advice on different fields of management. In 1791 Lavoisier, a tax collector before the revolution, wrote a report entitled *Summary of Various Works on Political Arithmetic.* The new republic urgently required this report because at the time taxation was based on property, the area of land that was under cultivation and the head of livestock on it.

Lavoisier attempted to calculate the quantity of land that was being worked in France, compiling data on the annual consumption of food and alcohol in both cities and villages and calculating the quantity of land required to produce this quantity of food and drink. Thanks to his study it is known that in 1790 France had 25 million inhabitants, 8 million of whom lived in cities and another 8 million were involved in wine production. Lavoisier lobbied for the creation of a national statistics office to create regular records on agriculture, commerce and population levels. Such was his faith in statistics that he believed their use would herald the end of political economy, making it possible to solve problems without compromise. (Nevertheless, in 1794 Lavoisier, the chemist, economist and wealthy aristocrat of the *Ancien Régime,* was guillotined.)

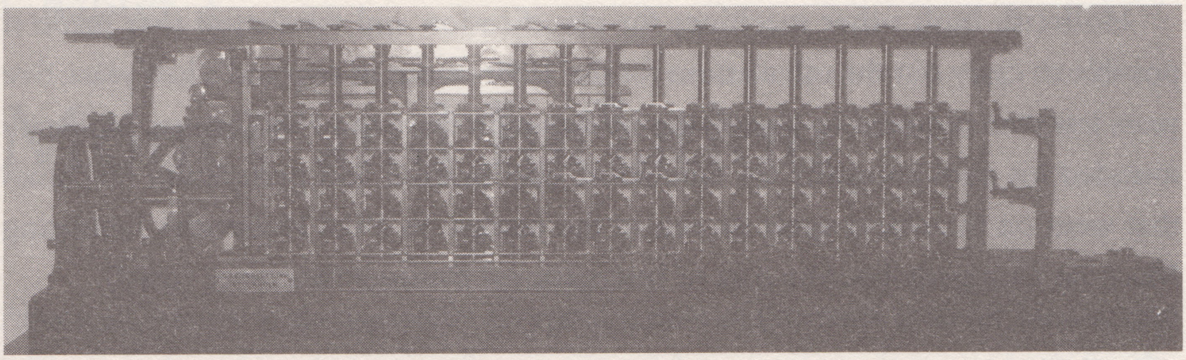
Another work of great importance was Thomas Malthus' *An Essay on the Principle of Population*, written towards the end of the 18th century and which would exert a powerful influence on many sociologists and economists. The work stated that since returns are diminishing and land is fixed (not infinite), historic trends indicate that food production grows arithmetically (1, 2, 3, 4…) while population levels grow geometrically (1, 2, 4, 8…). Hence as the population repeatedly doubles, it is as if the world were being divided in half over and over again, with the quantity of food available to satisfy the needs of the population decreasing until reaching a level below that required to sustain life, causing mortality to increase or salaries to become fixed at a minimum level of subsistence.

Up to 1799, 21 volumes of the *Statistical Account of Scotland* had been published, in which the politician Sir John Sinclair compiled an enormous collection of statistical data extracted from parish records, including annual summaries of revenue and expenditure by household and the types of activities through which the revenue was generated.

Towards the end of the 19th century the interest for quantifying economic and social data was growing. Government administrations gathered data for numeric political management, leading to cyclical population censuses and the development of mathematical models that made it possible to draw conclusions and prepare intelligible and usable reports for politicians, businessmen and researchers.

Demographic tables: the second birth of statistics

In 1839 the doctor William Farr applied statistical methods to health in the UK. He worked as a 'compiler of abstracts' at the General Register Office in London and in 1864 published the report *English Life Tables*, which included the first tables used by insurance actuaries to calculate risks for life insurance policies. The tables were based on records of births and deaths and census data and required complex numerical methods. The British government acquired a mechanical calculating machine with a printer, designed by the Swede Georg Scheutz, to carry out calculations.



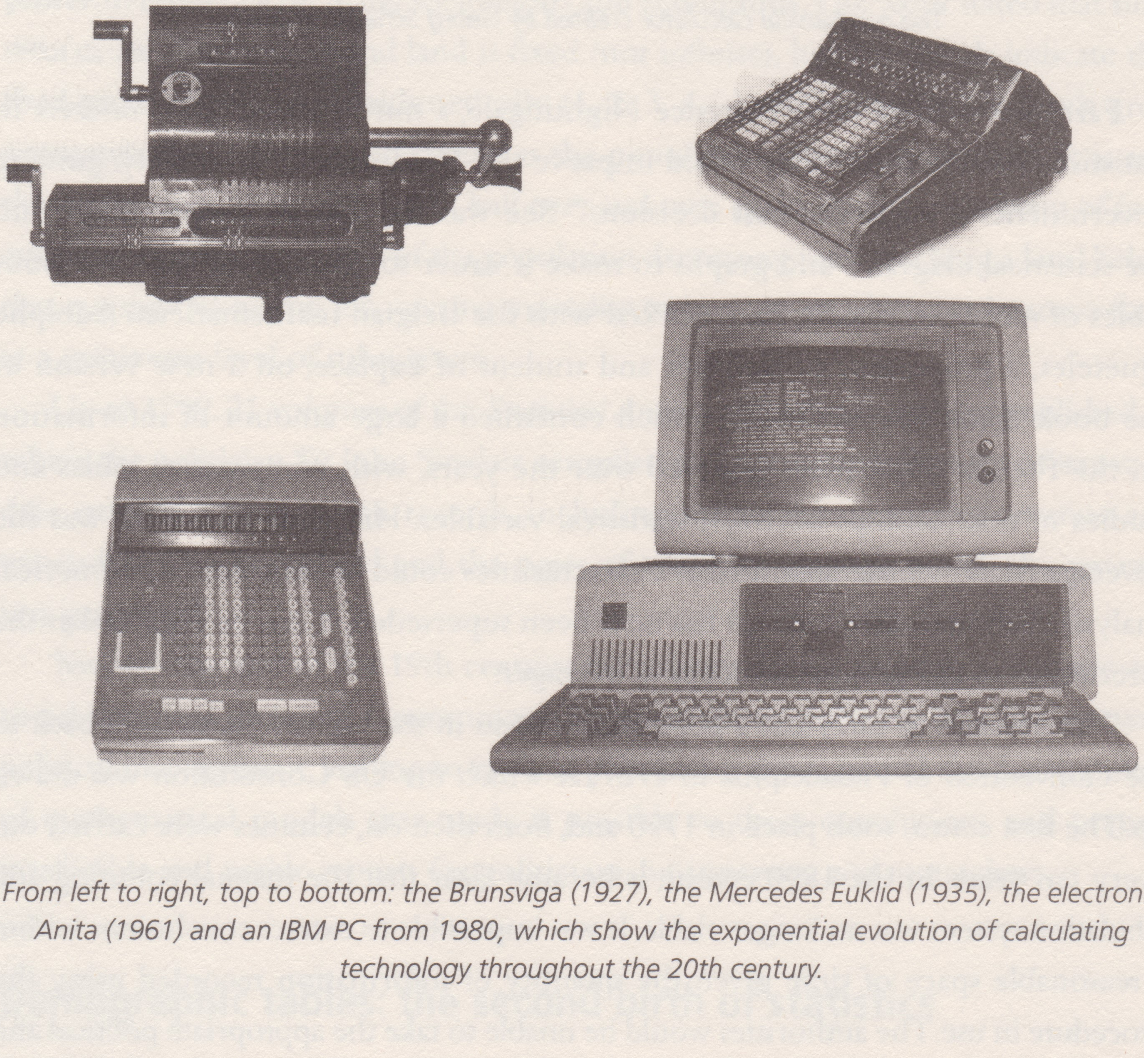
The mechanical calculator created by Georg Scheutz in 1856.

Farr also worked with Florence Nightingale, a nurse with a great interest in statistics. She believed "they could improve the conditions of humanity, guiding governments towards the best decisions." She was the first person to design and use statistical diagrams and graphs to make it easier to understand the enormous tables of numbers. Finally, Farr worked with the Belgian mathematician Adolphe Quetelet, a proponent of statistics and student of Laplace, on a new version of the book *Social Physics* (1835), which contained a large amount of information on the French population gathered over the years, with comparative tables and studies of relationships between statistical variables. His great invention was the 'Average Man', whose quantifiable characteristics could be subjected to numerical analysis. Although this concept has now been superseded, it can be regarded as the forerunner to the general concept of averages.

The decision to carry out a population census in the United States dates back to the Convention of Philadelphia in 1787, at which the US Constitution was drawn up. The first census took place in 1790 and, from then on, censuses were carried out every ten years. In the 1890 census, it became clear that the manual counting and tabulation system was no longer viable. It was impossible to extract conclusions within a reasonable space of time given the quantity of information recorded using the procedure in use. The authorities would be unable to take the appropriate political and economic decisions that would derive from the handling of this information before the time came to carry out the next census.

To overcome this situation, a competition was run to design a machine that would facilitate the quick and easy handling of census information. It was won by a punched-card machine designed by the engineer Hermann Hollerith. It used an electronic system to count the number of perforations made on the card, which indicated the value of each of the statistical variables collected on the census form. The perforations of each card had been made by an operator using the census forms.

Hollerith built other machines for classification and tabulation that made it possible to carry out the calculations for the 1890 census in two and a half years, five years less than required to classify the information from the 1880 census. Hollerith founded his own company for tabulation, classification and perforation machines that went on to achieve great success in the market. In 1914 Hollerith's Tabulating Machine Company became part of International Business Machines (IBM).



From left to right, top to bottom: the Brunsviga (1927), the Mercedes Euklid (1935), the electronic Anita (1961) and an IBM PC from 1980, which show the exponential evolution of calculating technology throughout the 20th century.

Between 1900 and 1935, the technology evolved from the Brunsviga arithmometer to the Mercedes Euklid, which implemented the four arithmetic operations and had a precision of 16 digits. The 1960s saw the introduction of electromechanical machines and finally the electronic semiconductor-based machines that form the basis of computing today. Programming languages were devised to control computers specifically designed for managing businesses, in addition to database management software such as SAP and DB-2, designed to handle the enormous volumes of data needed for the management of large organisations.

Chapter 2

Money and Inflation

Above all, money is the generally accepted means of payment used for buying and selling goods and services, as well as in financial transactions. It serves both for the purposes of exchange and for measuring the value of goods, wealth and assets. Hence, money is a unit of account for measuring the value of things as well as an instrument of payment. It has a tangible physical form (metal coins and banknotes) but it can also be unit for measuring the available balance in our bank account or a credit card. The latter is a plastic card whose chip or a magnetic strip records all the operations requested by the cardholder using an ATM or over the Internet, using access codes and cards with tables of random numbers.

A brief history of money from commodity money to fiat money

Money has evolved throughout history. To begin with, trade was based on the exchange of goods, on bartering merchandise. The idea was to exchange the surpluses of one community for the surpluses generated by another. Soon though, certain goods, such as units of livestock came to be used as reference values for bartering. Hence, for example, twenty amphorae filled with olives might have the value of a sheep, or one hundred amphorae filled with wine might have the value of an ox. Amphorae filled with water (with a capacity of between 25 and 30 litres), referred to as 'talents', were used as measures of weight and, later on, as monetary units too.

The adoption of units of livestock as a currency for payment led to these animal-currencies being represented on stones, clay tablets and, finally, on metal coins. Subsequent evolution led to metal coins minted from precious metals, such that the value of the coin lay in the metal it contained (known as commodity money).

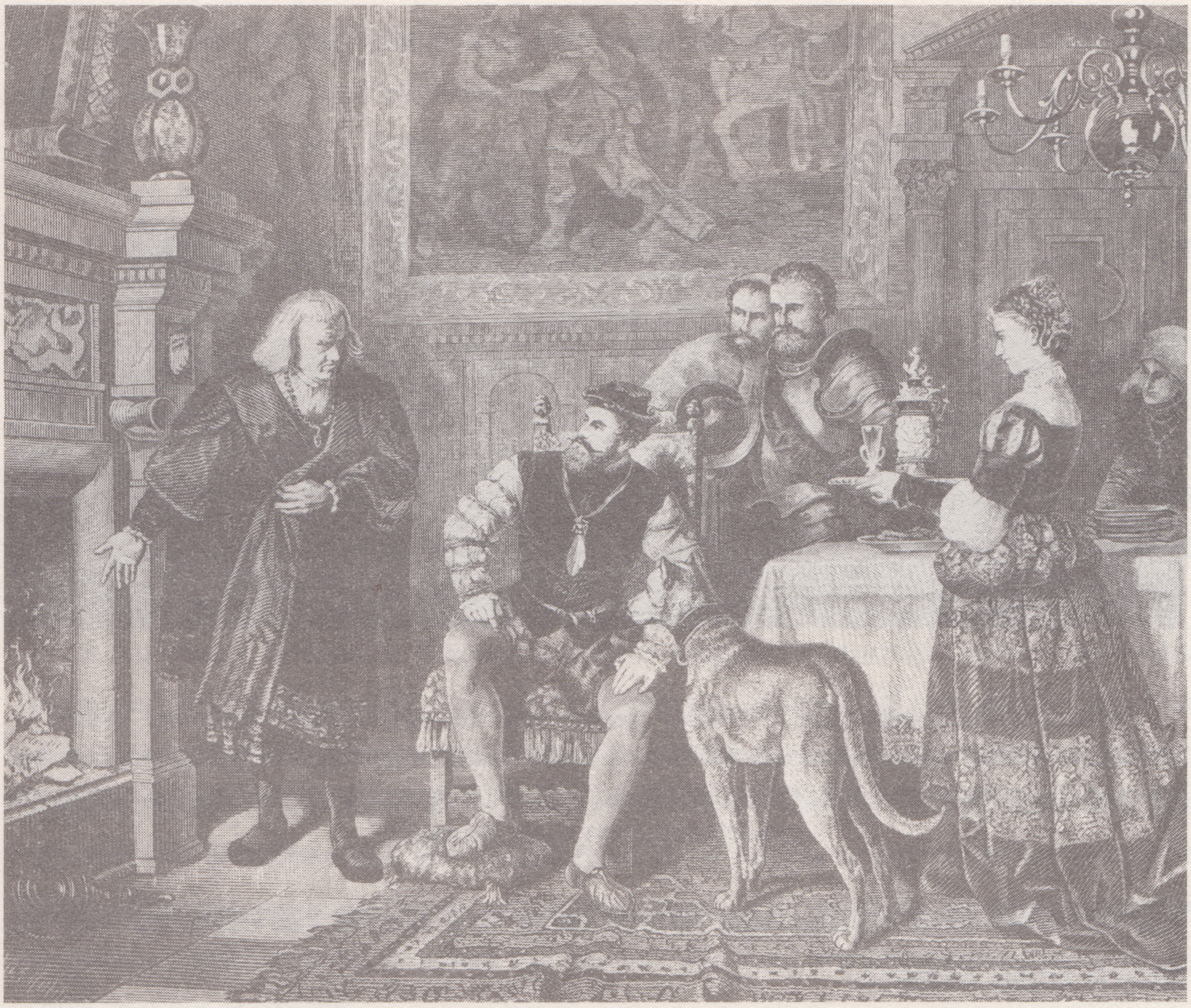
The next step was defined by a reduction in value of the minting metal, such that the intrinsic value of the money was much lower than the value it represented. This stage of so-called symbolic or 'fiat' money would culminate in the appearance of paper currency. However, as bronze, silver and gold coins continued to be minted, the merchants of the Middle Ages would ask metalsmiths to weigh the most valuable coins and determine their real value. For their part, the smiths would issue certificates for the coins deposited in their custody. Merchants soon came to regard these certificates as safer and more convenient, and the innovation of deposit certificates gave rise to not only the banknote but also the banking trade.

In antiquity, the value of coins was directly related to the material from which they were manufactured, as was the case with this bronze Roman sestertius.

Banks had deposits that were rarely called up by their owners and the smiths-cum-bankers soon came to realise that by setting aside just a small quantity of coins to repay their clients' infrequent demands, the remainder of these deposits could be used for loans, provided they maintained a certain 'liquidity coefficient'. This was the birth of banking money: if an account holder deposits money in a bank, and the bank lends the money to a third party, the money is in three different hands: the account holder, the bank and the borrower. Therefore the real money supply is the sum of currency in public circulation and bank deposits.

Monetary standards, fiat value and units of account

The monetary standard is the metal that guarantees the convertibility of a metal coin or a banknote upon presentation to the issuing bank. Each monetary system is based on a given metal: gold, silver or both (bimetallic) currency. Hence, governments and banks that issue paper currency guaranteed the automatic conversion into gold or silver of the issued bank notes, depending on whether the monetary system was based on the gold standard, or a bimetallic standard, governed by a



The Fugger family founded one of the main commercial and financial entities of the 15th and 16th centuries. In the illustration, King Charles V of Spain listens to Jakob Fugger, who financed his election as Holy Roman Emperor.

specific gold–silver ratio. Fiat money is based on public confidence in the issuing bank, which guarantees its automatic conversion into the precious metal of the monetary standard.

The gold standard lasted until World War I. In the aftermath of the war, it was replaced by the gold exchange standard, based on two strong currencies that guaranteed their exchange into gold: the US dollar and pound sterling. As all of the world's countries made payments for international trade in these strong currencies, they progressively came to be used as the basis for their reserves, until in 1971 the United States abolished the convertibility of the dollar into gold, which represented the definitive withdrawal of the gold standard. The standard was replaced by the fiat standard, which is based on confidence in strong currencies that form the reserves of the world's central banks.

The Mount Washington Hotel, in Bretton Woods, New Hampshire, where the United Nations Monetary and Financial Conference took place in 1944. This established a series of agreements that replaced those made after World War I and lasted until 1971, when the United States abolished the convertibility of the dollar into gold.

Currencies and exchange rates: why are some currencies worth more than others?

All money issued by a foreign central bank is denominated in a currency and constitutes the medium of payment used in international trade. Normally, international trade transactions are paid in the currency of the exporting country, although payments in strong currencies, such as the US dollar and the euro tend to dominate. The value for which one currency is exchanged for another is referred to as the exchange rate. Currencies are convertible if they can be exchanged with other currencies in line with the values determined by the free behaviour of the foreign exchange market, rather when the exchange rate that affects the currency is flexible.

Hence, on the foreign exchange markets, the dollar ($) can trade at a given value against the euro (€): $1 = €0.69 (which is the same as saying that $1.449 is needed to purchase €1); on the other hand, the exchange rates for pound sterling (£) may correspond to €1.11 and $1.61, and the rate for 100 Japanese yen (¥) to £0.0067, €0.0075 and $0.0108. This means that if we have euros, in order to purchase a machine in the United States that costs $10,000, we would have to purchase dollars to the value of €6,900; if we had pounds sterling, we would need to pay: 10,000/1.61 = £6,211.18.

When the currency has a fixed exchange rate, its government has established a permanent and constant exchange rate relating it to the other currencies, with little margin for variation. Convertible currencies that are freely traded on the foreign exchange markets exhibit bands of oscillation with respect to other currencies, but it is desirable that these do not exceed 1%, since this may affect the stability of exchange rates.

The determination of exchange rates, either floating or fixed, strongly influences the competitiveness of countries. Acute domestic inflation makes the products of a country expensive and reduces its possibilities for export, which also reduces its level of activity, with all the negative affects this can have on employment. This is why countries with high rates of inflation are forced to devalue their currencies with respect to others, in order to increase their international competitiveness.

A currency is said to depreciate with respect to another when the value required to purchase the currency increases. Hence, if the dollar were to depreciate with respect to the euro, this would mean that to purchase €1 we would need, for example, \$1.90 instead of \$1.449 as previously, meaning that in order to purchase €1 we would now have to pay a higher price since the dollar has lost value with respect to the euro. It is possible for a currency to depreciate with respect to some currencies while at the same time appreciating with respect to others.

Furthermore, the exchange rate between the currencies of two countries also influences changes in their price indices. Hence, for example, if Country 1 registers a greater increase in prices p_1 than the increase in prices p_2 registered in Country 2, the exchange rate between both currencies will vary in line with the change in the price indices for both countries.

Hence, if the exchange rate between the currencies of these countries is E, it will vary in line with the relative changes in prices, p_1 and p_2, hence:

$$\text{New exchange rate} = E \cdot \frac{p_1}{p_2}.$$

If the prices of Country 1 increase more than those of Country 2, its exchange rate will increase, meaning that it will tend to devalue. On the other hand, if the prices of Country 2 register a higher relative increase, the currency of Country 1 will tend to appreciate with respect to that of Country 2.

PURCHASING POWER PARITY AND THE VALUE OF CURRENCIES

To explain why some currencies are worth more than others, it is necessary to introduce another concept: the purchasing power parity of one currency with respect to another. The idea is that for $1 it is possible to purchase a basket of specified products; we must now investigate if it is possible to purchase more or less products using another currency at the current rate of exchange. Hence, if, for example, it is possible to purchase the same products in the basket of products purchased for $1 for 69 pence. Is it possible to purchase one hundred times the products of the $1 dollar basket for £69 that could also be purchased for $100? It is clear that if the dollar is devalued by 10.4% with respect to the pound, in this new situation, only 62.5 pence would be required to purchase the $1 basket, instead of the 69p previously required.

For example, if we purchase a camera in pounds for a price of £150 in the United States at an exchange rate of $1 = £0.69, the camera would cost £/$0.69 · $150 = £103.5.

If the dollar is devalued by 10.4% with respect to the pound, causing the exchange rate to drop to $1 = £0.625 (or £1 = $1.6), the camera purchased in pounds would become cheaper: £/$0.625 · $150 = £93.75.

However, it might also be the case that the euro loses its value with respect to the dollar (or the dollar appreciates with respect to the euro), meaning it is now only necessary to pay $1.25 for £1 ($1 = £0.80). In this case, the price of the camera increases for the British buyer: £/$0.80 · $150 = £120.

Hence, when a country devalues its currency with respect to that of another country, its exportable products become cheaper for the foreign country, and conversely, imports from the foreign country become more expensive. In this case, it is said that the country's purchasing power parity has deteriorated.

The real terms of trade of a country are defined by the quotient of the average price of its exports and the average price of imported products, hence:

$$\text{Real terms of trade} = \frac{\text{Average price of exports}}{\text{Average price of imports}}.$$

The higher a country's terms of trade, the greater its advantages in international trade, since this means it sells products at high prices in exchange for many more products at low prices. In international trade, all countries seek to have a comparative advantage, meaning that they try to export and import certain products that give them favourable terms of trade.

The exchange rates of the various countries are the result of the transactions carried out on the financial markets depending on the international payment requirements of the various agents: banks, central banks, multinational companies, financial institutions (e.g. investment funds, pension funds, insurance companies). The currencies in highest demand are those most widely used for international payments and are often those of the countries with the strongest economies. However, the demand for currencies is also influenced by the interest rates paid in the country of origin, in addition to expectations of the future price of the currencies. Some currencies are also influenced by the interventions of central banks to keep exchange rates undervalued and hence favour exports. This is the case with the United States, China and Japan, which keep their currencies devalued with respect to the euro in order to favour their international trade.

The simplest and most practical algorithms

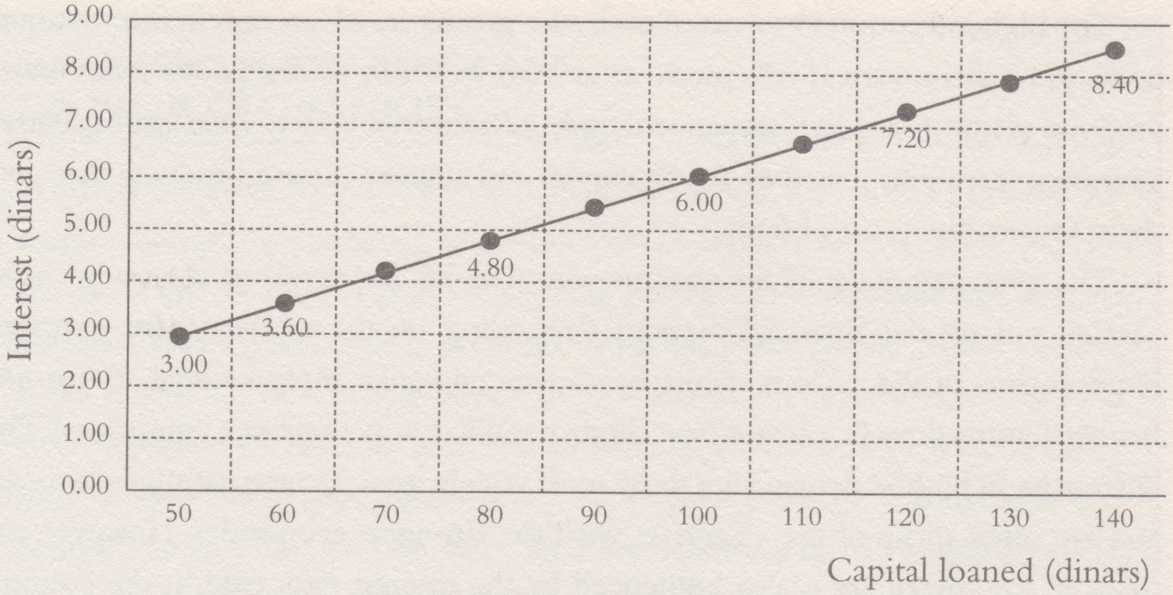
Let us now consider the simplest rules of mercantile arithmetic developed from the Renaissance until the end of the 20th century. The first of these is the direct rule of three which deals with problems in which two variables are directly proportional. (If one increases, the other also increases; if it decreases, the other decreases too). For example, if a moneylender earns 3 dinars on a loan of 50 dinars, how much would they earn on a loan of 120 dinars?

$$\begin{array}{rcl} 50 \text{ dinars} & \text{———————} & 3 \text{ dinars} \\ 120 \text{ dinars} & \text{———————} & x \text{ dinars.} \end{array}$$

The problem is solved as follows:

$$50/120 = 3/x$$
$$50x = 120 \cdot 3$$
$$50x/50 = 120 \cdot 3/50$$

$$x = 120 \cdot 3/50 = 7.20 \text{ dinars.}$$



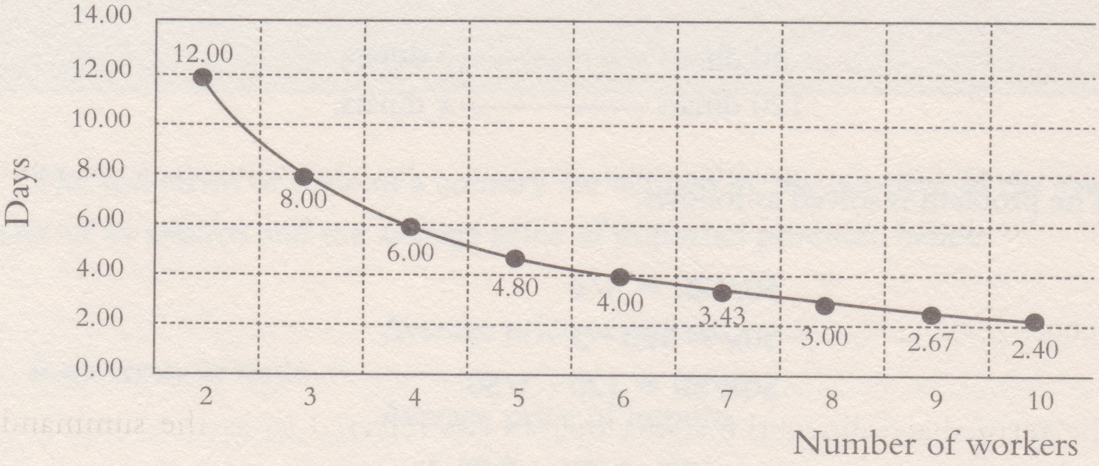
Direct rule of three.

On the other hand, the inverse rule of three sets solves problems in which two variables are inversely proportional (if one increases, the other decreases), such as in the following example. Two masons can build a wall in 12 days, how many days would it take five workers?

2 people ——————— 12 days
5 people ——————— x days.

The problem is solved as follows:

$$2/5 = x/12$$
$$2 \cdot 12 = 5x$$
$$5x/5 = 2 \cdot 12/5$$
$$x = 2 \cdot 12/5 = 4.80 \text{ days} = 4 \text{ days, } 19 \text{ hours, } 12 \text{ minutes.}$$



Inverse rule of three.

Finally, a composite rule of three tackles problems of the following type: if 40 workers working eight-hour days paint 320 m of fencing in 10 days, how many days would 55 painters take to paint 440 m of the same fencing working six hours a day?

Painters	Hours	Metres	Days
40	8	320	10
55	6	440	x
	Inverse	Inverse	Direct

The solution to the problem is as follows:

$$10/x = 55/40 \cdot 6/8 \cdot 320/440$$
$$x = 10 \cdot 440/320 \cdot 8/6 \cdot 40/55 = 13.3 \text{ days} = 13 \text{ days, } 8 \text{ hours.}$$

The sigma operator

The sigma operator (Σ, the capital Greek letter sigma) is commonly used in mathematical formulations in economic theory and represents summation. It indicates that the operation of addition should be carried out. Hence, to indicate $x_1 + x_2 + x_3 + x_4$ we can use the expression $\sum_{i=1}^{4} x_i$.

The symbol Σ, when placed before x_i means that we need to add all the values of x_i and the numbers that are above and below indicate the limits of the summation, or rather the lower and upper bounds of the subscript indices that must appear in the summation being carried out.

The sum $\sum_{k=3}^{6} x_k$ is $x_3 + x_4 + x_5 + x_6$ and the sum $\sum_{j=m}^{n} x_j$ is $x_m + x_{m+1} + \ldots + x_{n-1} + x_n$.

Subscript indices may only take integer values. The total summation does not change if we modify the letter of the subscript index.

Hence, $\sum_{i=1}^{m} x_i = \sum_{j=1}^{m} x_j = \sum_{k=1}^{m} x_k$.

The term that follows the sigma operator is referred to as the summand. In $\sum_{k=1}^{m} x_k$, the summand is x_k.

EQUATIONS AND INEQUATIONS

An equation is a mathematical expression that includes an equality (equals sign), numbers or constants, and letters, referred to as unknowns or variables. The expression is only true for certain values of these unknowns. The simplest equation has only one unknown that is not squared or cubed. For example, $x + 12 = 25 - 3x$ is a first-degree equation, $12 + x^2 - 6x = 3$ is a second-degree equation, and $9 - 3x^2 - 6x^3 = -12$ is third degree.

In the 13th century, Leonardo of Pisa solved problems like this modern version. A silversmith has gold with a fineness (purity) of 0.975 (97.5%) and gold with a fineness of 0.750 (75%), and wishes to obtain a 2 kg ingot with a fineness of 0.900 (90%). How much gold of each type is required? The problem is solved as follows:

x kg ———— the weight of the 0.975 gold
$(2 - x)$ kg ———— the weight of the 0.750 gold

$x \cdot 0.975 + (2 - x) \cdot 0.750 = 2 \cdot 0.900$
$x \cdot 0.975 + 2 \cdot 0.750 - 0.750 \cdot x = 1.800$
$x \cdot 0.975 - 0.750 \cdot x = 1.800 - 2 \cdot 0.750$
$x \cdot 0.225 = 1.800 - 1.500$
$x \cdot 0.225 = 0.300$
$x = 0.300/0.225 = 4/3 = 1\ 1/3$ kg at 0.975
$(2 - x) = 2 - 1\ 1/3 = 2/3$ kg at 0.750.

Fibonacci also set and solved second-degree problems, such as the following. The surface of a rectangular field is 2,400 m^2. It is known that it is 20 m longer than it is wide. Calculate the dimensions of the field. Hence, the width (x) multiplied by the length ($x + 20$) gives 2,400 m^2. The standard form of writing a second-degree equation is: $ax^2 + bx + c = 0$ and the formula that allows us to calculate the unknown x is:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The sum of the first eight odd numbers can be written as follows:

$$\sum_{j=0}^{8}(1+2j) = (1+2\cdot 0)+(1+2\cdot 1)+(1+2\cdot 2)+(1+2\cdot 3)+(1+2\cdot 4)+(1+2\cdot 5)+$$
$$+(1+2\cdot 6)+(1+2\cdot 7)+(1+2\cdot 8) = 1+3+5+7+9+11+13+15+17.$$

In this case:

$x \cdot (x+20) = 2{,}400;\ x^2 + 20x = 2{,}400;\ x^2 + 20x - 2{,}400 = 0$

$$x = \frac{-20 \pm \sqrt{20^2 - 4 \cdot 1 \cdot (-2{,}400)}}{2 \cdot 1} = \frac{-20 \pm \sqrt{10{,}000}}{2} = \frac{-20 \pm 100}{2} = \left\langle \begin{array}{l} \frac{-20+100}{2} = 40 \text{ m} \\ \frac{-20-100}{2} = -60 \text{ is not possible} \end{array} \right.$$

$x + 20 \text{ m} = 40 + 20 = 60 \text{ m}.$

Hence, the solution obtained for the dimensions of the field is 40 m × 60 m.

Inequations, on the other hand, are similar to equations, but instead of having an equals sign (=) they have an inequality sign, of which there are four types:

$\leq$ less than or equal to

$<$ less than

$\geq$ greater than or equal to

$>$ greater than.

The single variable inequation $x - 7 \geq 13$ represents all the numbers which, when subtracted from 7 give a result greater than or equal to 13. Inequations are solved in a similar way to equations. For example:

$x - 7 \geq 13;\ x - 7 + 7 \geq 13 + 7;\ x \geq 20.$

The solution to this inequation is the set of all numbers greater than or equal to 20.

In some cases, equations and inequations behave differently, such as in the following example:

$$\text{equation} -\frac{x}{12} = 3;\ -x = 3 \cdot 12;\ -1 \cdot (-x) = 1 \cdot 36;\ x = -36.$$

$$\text{inequation} -\frac{x}{12} \geq 3;\ -x \geq 3 \cdot 12;\ -1 \cdot (-x) \leq (-1) \cdot 36;\ x \leq -36.$$

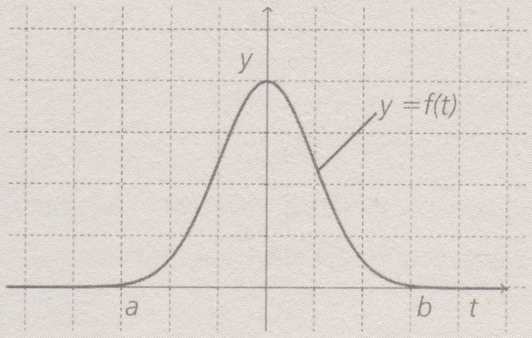
In this example, it is necessary to invert the inequality sign to solve the inequation. We can check this by noting that if 7 < 13, whereas: − 7 > − 13.

The sum $\sum_{j=2}^{5} 2^j$ gives $2^2 + 2^3 + 2^4 + 2^5 = 4 + 8 + 16 + 32.$

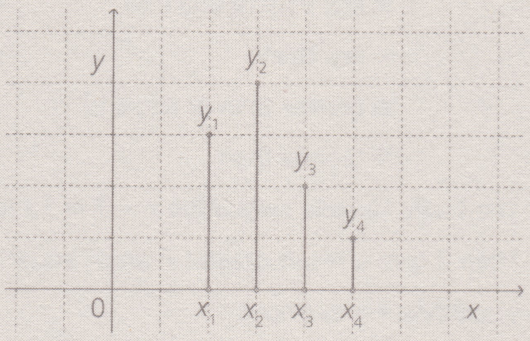
And the sum $\sum_{l=1}^{3} (l+1) \cdot 3^l = 2 \cdot 3^1 + 3 \cdot 3^2 + 4 \cdot 3^3 = 6 + 27 + 108.$

DISCRETE AND CONTINUOUS VARIABLES

In many modern mathematical concepts, a variable is defined by a discrete set (which means the variable can only take certain values and there is no value between them). This is indicated as $\{x_1, x_2, \dots x_n\}$. Between the values x_1 and x_2 the variable x has no values. There are other variables, much more commonly used, that define continuous sets (meaning that the variable can take all values, including integers, fractions and irrational numbers). One such example would be $\{0 \leq t \leq \infty\}$. It is often necessary to integrate ($\int f(x)\,dx$) to solve various problems related to functions defined in continuous sets, as in the case of the normal probability function or distribution. For discrete variables, the operation analogous to integration for continuous variables is summation (indicated using the Greek letter Σ or sigma).



Function $f(t)$ *of a continuous variable* t *defined by the set* $\{a \leq t \leq b\}$.



Function $y(x)$ *of a discrete variable* x *defined by the set* $\{x_1, x_2, x_3, x_4\}$.

A set of four objects can be named using letters (and numbers as subscript indices): x_1, x_2, x_3 and x_4. To speak of a set of n elements (n can vary depending on the problem) we use the notation $\{x_1, x_2 \dots, x_{n-1}, x_n\}$. Here, x_{n-1} indicates the number before x_n (the last term). To designate the element in the series that occupies position i, we write x_i. In this way, for example, we can indicate the prices of four products as p_1, p_2, p_3 and p_4, and the quantities required of each as q_1, q_2, q_3 and q_4.

One of the most important applications of Σ is in expressing mathematical series, such as the binomial series. The binomial law of probability is applied to the study of surveys with two possible responses (e.g. 'yes' and 'no') with a probability of each occurring being p and $q = (1 - p)$, values determined based on previously gathered data. The answers 'yes' and 'no' are two complementary events and the sum of their probabilities $p + q = 1$, hence $q = 1 - p$.

To find the probability of three consecutive 'yes', we must first calculate the probability that one of the following events occurs: no 'yes', one 'yes', two 'yes', or three 'yes', which is $P(0 < k < 3) = P(0) + P(1) + P(2) + P(3)$, whereas the formula written using the sigma operator is:

$$\sum_{k=0}^{3}\binom{4}{k}p^{k}q^{n-k}=\binom{4}{0}p^{0}q^{4}+\binom{4}{1}p^{1}q^{3}+\binom{4}{2}p^{2}q^{2}+\binom{4}{3}p^{3}q=q^{4}+4pq^{3}+6p^{2}q^{2}+4p^{3}q.$$

The function that makes it possible to calculate the probability of obtaining 'yes' between 0 and k times with n surveys is the sum of the probabilities up to $P(k)$. Using the sigma operator, this can be written as follows:

$$P(0)+P(1)+P(2)+...+P(k)=P(0\leq k\leq n)=$$

$$=\sum_{k=0}^{k}\binom{n}{k}p^{k}q^{n-k}=\binom{n}{0}p^{0}q^{n}+\binom{n}{1}p^{1}q^{n-1}+\binom{n}{2}p^{2}q^{n-2}+...+\binom{n}{k-1}p^{k-1}q^{1}+\binom{n}{k}p^{k}q^{0}.$$

The Σ operator is also applied to statistical functions, such as:

$$\left.\begin{array}{l}\sum x_{i}=a\cdot\sum y_{i}+b\cdot n\\ \sum x_{i}y_{i}=a\cdot\sum y_{i}^{2}+b\cdot\sum y_{i}\end{array}\right\},$$

which, in expanded form, is:

$$\left.\begin{array}{l}x_{1}+x_{2}+...+x_{n}=a\cdot(y_{1}+y_{2}+...+y_{n})+b\cdot n\\ x_{1}y_{1}+x_{2}y_{2}+...+x_{n}y_{n}=a\cdot(y_{1}^{2}+y_{2}^{2}+...+y_{n}^{2})+b\cdot(y_{1}+y_{2}+...+y_{n})\end{array}\right\}.$$

Similarly, it has an application in statistical formulae:

$$r=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i}-\overline{x})\cdot(y_{i}-\overline{y})}{\sigma_{x}\cdot\sigma_{y}}=\frac{\frac{1}{n}\left[(x_{1}-\overline{x})\cdot(y_{1}-\overline{y})+(x_{2}-\overline{x})\cdot(y_{2}-\overline{y})+...+(x_{n}-\overline{x})\cdot(y_{n}-\overline{y})\right]}{\sigma_{x}\cdot\sigma_{y}},$$

$$\sigma=\sqrt{\frac{(x_{1}-\overline{x})^{2}+(x_{2}-\overline{x})^{2}+...+(x_{n}-\overline{x})^{2}}{N}}=\sqrt{\frac{\sum_{k=1}^{n}(x_{k}-\overline{x})^{2}}{\sum_{i=1}^{n}f_{i}}},$$

$$\overline{x}=\frac{x_{1}f_{1}+x_{2}f_{2}+....+x_{n}f_{n}}{N}=\frac{\sum_{i=1}^{n}x_{i}f_{i}}{\sum_{i=1}^{n}f_{i}};\ N=\sum_{i=1}^{n}f_{i}=f_{1}+f_{2}+...+f_{n}.$$

How is inflation measured?

Inflation is the annual rise in the prices of goods and services, such that if the salaries or incomes of consumers do not change, their purchasing power deteriorates, meaning they can purchase less items for the same value. Inflation is measured by the variations in prices. The effects of high inflation are damaging to a country's economy, since they reduce the wage levels of consumers and families while also reducing economic competitiveness in terms of international trade; hence the requirement to keep it under control. Price stability is one of the conditions for keeping a country's economy healthy.

Variations in prices are measured using indices. The prices are in fact the average prices of a basket of goods and services, in which the average of various prices is calculated using the relative weights of the products that make up the basket used for calculating the average price. The calculation starts with a base year, which is given a value of 100, and to which the increases in the average weighted prices recorded in subsequent years are applied.

For example, if we take year zero as the base year with an index = 100 and record the following price increases for subsequent years, the evolution of the price index will be as follows:

	Δ annual in %	Price index	Formula for Calculation
Base Year	0.00	100.000	100 + 0
Year 1	1.25	101.250	100 + 1.250
Year 2	2.30	103.579	101.250 + (101.250) · 0.023
Year 3	1.50	105.132	103.579 + (103.579) · 0.015
Year 4	2.00	107.235	105.132 + (105.132) · 0.02
Year 5	3.00	110.452	107.235 + (107.235) · 0.03

Calculation of the price index.

A country's department of statistics prepares various price indices for various products, but the best known is the Retail Price Index (RPI), calculated based on a fixed basket of specific products normally used by consumers (food, clothing, housing, education, transport, etc.). Price indices are also calculated for other products in sectors such as industry, agriculture, construction and property.

The annual rate of inflation (in %) is calculated as follows:

$$\text{Rate of inflation} = \frac{\text{Index for this year} - \text{Index for previous year}}{\text{Index for previous year}} \cdot 100.$$

There are many theories to explain why prices rise. Some are explained algebraically, such as the equation of exchange developed by Irving Fisher. In line with this theory, the level of prices, P, of a country depends on the total volume of economic transactions (T), the total money in circulation (M) and the velocity of the circulation (V) of money (i.e. the number of times money changes hands).

$$P \cdot T = M \cdot V.$$

PRICE INDICES

To quantify changes in prices, production levels or values over a given period of time, we make use of indices, of which there are an almost infinite number of definitions.

If we wish to calculate general indicators for changes in prices, production levels or values for a given country, we must base them on a series of their values over time. The comparison is made with respect to a given year, let's say year 0. The production levels are calculated at constant prices, and the prices at constant production.

The most widely used indices for production levels and prices are the Laspeyres (Q_L, P_L), Paasche (Q_P, P_P), and Fisher (Q_F, P_F) indices.

$$\text{Laspeyres index: } Q_L = \frac{\sum p_0 q_t}{\sum p_0 q_0} \qquad P_L = \frac{\sum p_t q_o}{\sum p_0 q_0} \;;$$

$$\text{Paasche index: } Q_P = \frac{\sum p_t q_t}{\sum p_t q_0} \qquad P_P = \frac{\sum p_t q_t}{\sum p_0 q_t} \;;$$

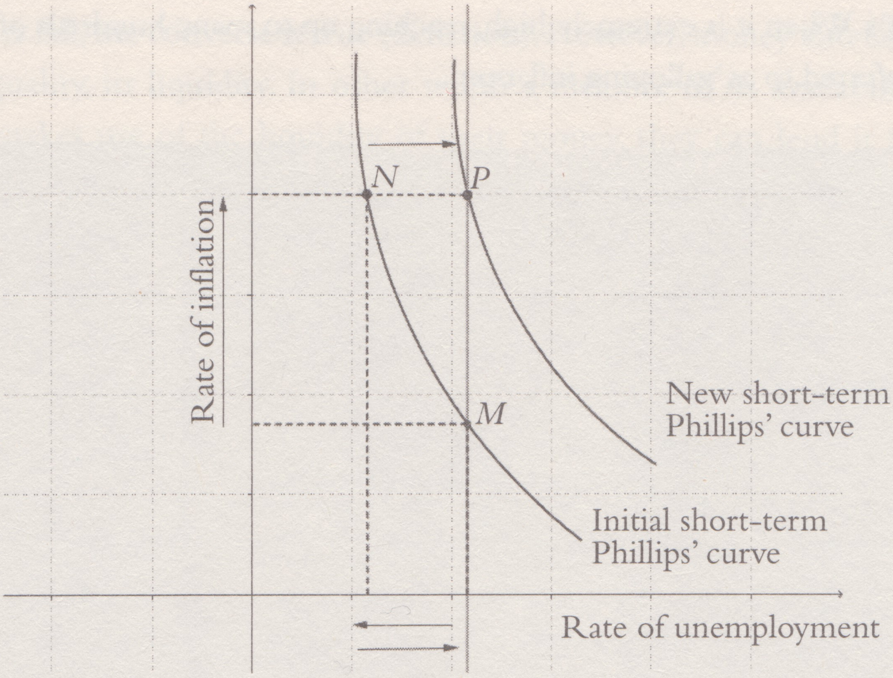
$$\text{Fisher index: } P_F = \sqrt{P_L \cdot P_P} \qquad Q_F = \sqrt{Q_L \cdot Q_P} \;,$$

where P = Prices and Q = Production levels, and $\sum p_0 q_t$ indicates the sum of the values of all goods produced by a country in year t at a constant price p_0 for year 0, and $\sum p_0 q_0$, the sum of the values of goods produced by a country in year 0, and the same for production values at a constant price. The Fisher index is the geometric mean of the Paasche and Laspeyres indices.

The US economist Irving Fisher (1867–1947). In addition to other contributions to the theory of economics, he developed the equation of exchange.

Solving this equation for the variable P, gives a new relationship: $P = M \cdot V / T$. Based on this, we can easily deduce that prices will increase if M, the publicly available money supply, increases and the other variables, V and T, remain constant, or if V, the velocity of circulation of money, increases. Similarly, prices will decrease when the amount of money or velocity of circulation decreases. On the other hand, price variations are inversely proportional to the volume of transactions: If the supply of products increases and the amount and velocity of circulation of money remain constant, prices will fall, and if transactions decrease (production falls), prices will rise.

There are other theories to explain the behaviour of inflation, such as the curve devised by A.W. Phillips. This related the rate of unemployment to the rate of inflation using statistical data for the United Kingdom, preparing a graphical representation with the rate of inflation on the vertical axis and the rate of unemployment on the horizontal. Phillips' curve is decreasing and behaves in a similar way to a demand curve – as prices rise, demand falls. The result was that he obtained an inverse relationship between unemployment and inflation. Put another way the unemployment rate falls with inflation, and conversely, a policy that seeks to curtail prices will necessarily result in an increase in unemployment.



A Phillips' curve.

Price inflation can occur for many reasons. An excessive increase in demand can cause prices to rise if production (supply) remains unchanged. This is what is known as demand-pull inflation. Rising wages, as a consequence of a reduction in unemployment, as well as higher costs of raw materials, can also cause inflation. Prices rise even though a certain rate of unemployment is maintained, rising faster when there is more economic expansion. During the last century, prices have always risen, even when there has been a recession. In times of recession, cost-push inflation can also occur, where prices increase in spite of the fact that unemployment is also rising.

In the event of severe economic recessions, situations can also arise where prices fall, in other words deflation (negative inflation). Deflation occurs when the money supply and the speed of circulation of employment reduce as a result of the increase in unemployment and the contraction of credit, and, at the same time, there is excess production that pushes companies to lower their prices in order to be able to sell surplus stock.

Another situation is hyperinflation, which occurs when governments decide to inject large quantities of money into circulation (in other words, set the money printing presses to full speed) to cover their expenses or inject liquidity to consumers, without the country's productivity structure changing and with productive activity remaining stagnant or in recession. Hyperinflation often reaches two or more digits, or more than 1,000%. When it is extremely high, reaching up to many hundreds of percent a year, it is referred to as 'galloping inflation'.

148E 048486
Reichsbanknote
Eintausend Mark
zahlt die Reichsbankhauptkasse in Berlin gegen diese Banknote dem Einlieferer
Berlin, den 15. Dezember 1922
Reichsbankdirektorium
1000

In the 1920s, when Germany suffered one of the worst periods of hyperinflation in history, a 1,000 mark note was converted into a 1000 million mark note.

When, during a period of economic recession, such as during the oil crisis at the start of the 1970s, a large increase in prices is recorded accompanied by an increase in unemployment, the phenomenon referred to as stagflation occurs.

Interest rates: the cost of money

In principle, the nominal value of money does not change – a dollar is a dollar, a euro is a euro, and a pound is a pound. Money is used as a unit of account to measure the value of things (goods and services, houses, land, wealth). It is also a generally accepted medium of payment, and so acquires a real form, such as metal coins and bank notes. However, money can also be used to purchase goods and services, which often vary in price, such that over the course of time, rising prices mean we can buy fewer and fewer items for the same quantity of money, as it loses its purchasing power. Put another way, inflation means we need more money to buy the same things.

Money derives its value from the valuation of the goods that can be purchased with it. When there is inflation, it is said that the value of money falls, whereas when prices fall, the converse is true (deflation). However, money also has another inherent quality, its liquidity. In other words a measure of its availability. When someone makes use of the liquidity of their money, they can lend it to a third party who needs it to acquire the goods and services they require, and who is willing to pay an additional price for its availability for a given period of time in addition to paying back the money itself. The price a third party is willing to pay for money that has been loaned is called the rate of interest. Hence, the interest rate is the price of money.

The interest rate is defined as a percentage to be paid for a sum of capital that is loaned. Hence, if we obtain a loan for £1,000 over three years at an annual interest rate of 5%, every year we will have to pay £50 in interest and at the end of the period, when the loan has been repaid, we will have repaid a total of £1,150.

A person may require a given sum to purchase an item, or an entrepreneur may need another sum to buy a machine, with neither having the money required to meet their needs. To do so, they request a bank loan. However in exchange for being able to have the money, they must repay it within a specified period and pay a price for the availability of this money. The price they will pay for the money that is loaned is the interest rate. On the other hand, another citizen with free money relinquishes the liquidity of the money, having deposited it in a bank, also in exchange for a return that is paid at an interest rate. As such, the money has a price both for the bank, which must obtain it from its customers by paying a price for it, and for the borrower, who must return the loan with the addition of interest.

The interest rate is also the price paid as a return on certain financial assets. Hence, if a government issues public debt, bonds, treasury bills, etc. to obtain funds, in addition to the amortisation terms of the securities, it pays a specific interest rate. For example, the state can issue bonds at an interest rate of 3.5% or short-term bills for 3 or 6 months at 1% and 1.5% respectively. Private companies can also issue fixed return securities (also called bonds) on the capital market with specific interest rates. These securities in the form of small loans allow companies to obtain money from the capital markets, but this has a cost – the interest rate.

The different interest rates of central banks

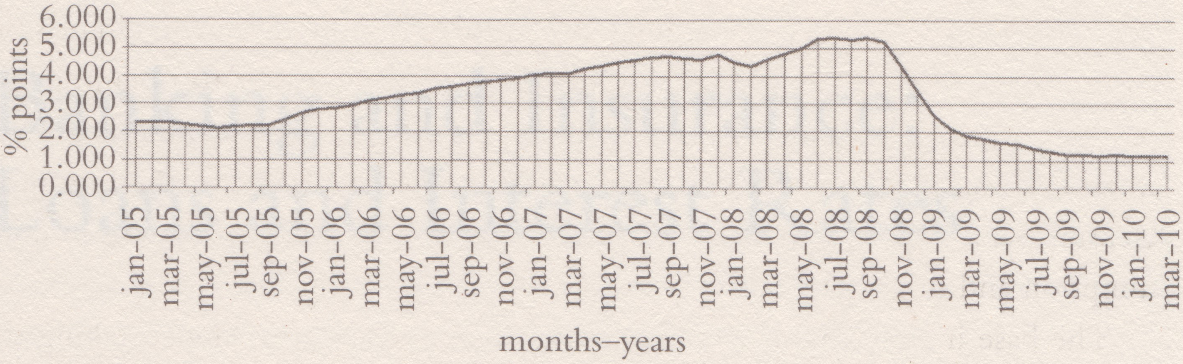
The central banks of each country have a base interest rate, which is the official interest rate, to control the amount of money in circulation and, in a direct way, inflation as well. If they decide to lower the interest rate, they will increase the money supply in circulation, making credit cheaper and incentivising private consumption and business investment. When there is more money in circulation, consumers will have more access to liquidity, it will be easier to obtain credit, demand will increase and prices will rise. On the other hand, if economic growth is excessive and inflation soars, central banks (the Federal Reserve, the European Central Bank, etc.) will tend to raise base interest rates in order to rein in consumption and rising prices.

The base interest rate varies by country or monetary area. Hence, taking the figures for January 2010, the US Federal Reserve System (FED) had its rate fixed at 0.25%, the European Central Bank at 1%, the Bank of England at 0.50%, the Bank of Japan at 0.10%, the Bank of Canada at 0.25%, the Reserve Bank of Australia at 3.75%, the Swiss National Bank at 0.25%, etc. The official interest rate is that which is used as a reference and starting point for the auctions of money between financial entities to fix the interbank interest rates, or rather the interest rate (price of money) at which banks lend money among themselves.

Central Bank	Interest Rate
European Central Bank	1%
Bank of Canada	0.25%
Bank of England	0.50%
Bank of Japan	0.10%
Reserve Bank of Australia	3.75%
Swiss National Bank	0.25%
United States Federal Reserve	0.25%

Official interest rates (January 2010).

These interbank interest rates are often slightly higher than the official interest rates. Euribor (Euro InterBank Offered Rate) is the interbank interest rate that is used in the Eurozone and is calculated based on the interbank rate used in open market operations by the main European banks, forming the base interest rate for the majority of loans. For consumer credit or mortgages, banks often use the Euribor (or Bank of England or FED interest rate if the loan is in the UK or the United States) as the base interest rate plus a number of points.



Change in the Euribor between 2005 and 2010.

In fact, there is a Euribor rate for each of the different credit operations: 1 month, 3 months, 6 months, 12 months... Similarly, the interbank base interest rates for the FED, the Bank of England and the Bank of Japan also differ depending on whether they are for 1 day, 3 months or 1 year. The Euribor rate is calculated every day before 11 o'clock in the morning, based on the rates applied by the 64 largest European banks: extreme values are eliminated (15% higher and 15% lower) and the arithmetic mean is calculated. The one-day interbank rate, Eonia (Euro OverNight Index Average), is calculated by the European Central Bank based on the daily interbank rates used by 48 European banks.

Chapter 3

Banking and Insurance: Loans and Interest Rates

The impact of mathematics on the world of banking is most strongly felt when it comes to calculating the instalments of loans and mortgages. The use of functions and progressions are fundamental to financial decisions, as we shall see on the following pages.

Simple interest and compound interest

First of all, we need to define certain basic concepts, such as capital, return and interest rates. In economics, capital is a factor of production. It is the value provided by the owning entrepreneur who invests in facilities and machinery, that is to say in so-called capital goods or production investment. In the world of finance, it is associated with the quantity of money possessed by someone, which is deposited in a bank and used for the purchase of bonds, variable and fixed-rate securities, and other types of financial assets. In other words, it is loaned to a third party for a given period in return for a sum of money, which is calculated applying a specific periodical percentage to the capital loaned (daily, monthly, annually, etc.), referred to, as we saw in Chapter 2, as the interest rate.

Hence, when someone requests a specific amount of capital C_0 from a bank, for example, for a period of three years ($n = 3$) and at an annual interest rate of 6% (nominal interest rate $i = 6\%$), at the end of the three-year period they will have repaid the capital and three times 6% of the capital, which is the return or value of the interest resulting from applying the agreed rate. For example, if $C_0 = 1{,}000$, the duration of the loan $n = 3$ years, and the interest rate $i = 6\%$, at the end of the three years, the person will have repaid 1,000 units of money, plus $3 \cdot (6/100) \cdot 1{,}000 = 180$, or rather, the capital plus a return (interest) of 60 monetary units per year (1,180 in total).

When the same quantity of interest is paid each year, the interest rate is said to be simple, and the final capital C_n that must be returned at the end of the period is made up of the initial capital plus the interest accrued, and will be:

$$C_n = C_0 + n \cdot i \cdot C_0 = C_0 (1 + n \cdot i).$$

This is the formula for simple interest, where C_0 is the capital that has been loaned; *i*, is the interest rate applied as a percentage of the capital (divided by 100); *n*, the number of time periods for the loan for which the interest is applicable; C_n, the total initial capital plus accrued interest; and $n \cdot i \cdot C_0$,the total interest accrued during the period for which the loan is valid.

The converse is also valid, whereby the customer of the bank deposits a sum of capital as a fixed-term deposit. The bank will pay a periodic interest rate on the deposit, for example at the end of each year. In this case, the bank will pay the accrued interest into the customer's current account, and the payment can be made annually, quarterly, twice a year or monthly, as determined in the contract for the fixed-term deposit.

The interest rate that is applied can be an annual rate payable, for example, once a year, month or quarter. In this case, the instalment of the interest paid into the customer's current-account will be the annual interest rate, or divided by twelve or four where applicable. However, the interest rate can also be monthly, quarterly, or any other fraction of the year. In this case, the same formula is applied, expressing the time *n* in months, quarters or whatever period is appropriate.

It may also be the case that the customer wishes to reinvest their interest returns on their initial deposit at the same interest rate. In this case, the new type of interest that is applied is referred to as compound interest. Hence, continuing with the previous example, at the end of the first year the customer will reinvest the final capital of 1,060 monetary units. At the end of the second year, they will have capital of 1,123.60, since in addition to the 120 monetary units of interest, they will also have 6% of the 60 units earned during the first year, which represents 3.6 units. Finally, at the end of the third year, they will have a final capital of 1,191.02, or rather, an overall profitability of 19.10% for the total period, an additional 1.1 percentage points than they would have obtained with simple interest.

The type of interest applied to a loan or as a return on capital can be monthly, quarterly or annual. Hence, a nominal interest rate of 12% applied monthly with a 1% cumulative monthly rate, is not the same as a 12% nominal rate applied annually. This leads us to the definition of the annual equivalent rate. A loan with an annual interest rate *i*, applied in *n* periods of the year (e.g. monthly) will have the following

GENERAL FORMULA FOR COMPOUND INTEREST

Generalising for a period of n years, the general formula for compound interest for a deposit or loan with an initial sum of capital C_0 is obtained as follows. In the first year $n = 1$ it will generate interest of $C_0 \cdot i$. In the second period, $n = 2$, the initial capital will increase $C_1 = C_0 + C_0 \cdot i = C_0 \cdot (1 + i)$. And so on, until the final year, n (hence i for each).

$n = 0$, C_0,
$n = 1$; $C_1 = C_0 + C_0 \cdot i = C_0 \cdot (1 + i)$,
$n = 2$; $C_2 = C_1 + C_1 \cdot i = C_0 \cdot (1 + i) + C_0 \cdot (1 + i) \cdot i = C_0 \cdot (1 + i) \cdot (1 + i) = C_0 \cdot (1 + i)^2$,
$n = 3$; $C_3 = C_2 + C_2 \cdot i = C_0 \cdot (1 + i)^2 + C_0 \cdot (1 + i)^2 \cdot i = C_0 \cdot (1 + i)^2 \cdot (1 + i)^2 = C_0 \cdot (1 + i)^3$,
……………
$n = n$; $C_n = C_0 \cdot (1 + i)^n$.

This gives the following formula for compound interest: $C_n = C_0 \cdot (1 + i)^n$. This formula allows us to deduce the interest rate i or the number of periods n, given the values of the other variables.

$$C_n = C_0 \cdot (1+i)^n; \ \frac{C_n}{C_0} = (1+n)^n;$$

$$1+i = \sqrt[n]{\frac{C_n}{C_0}};$$

$$i = \sqrt[n]{\frac{C_n}{C_0}} - 1.$$

Furthermore, writing the formula $C_n = C_0 \cdot (1 + i)^n$ in logarithmic notation gives:

$$\log C_n = \log C_0 + n \cdot \log(1+i);$$

$$n \cdot \log(1+i) = \log C_n - \log C_0;$$

$$n = \frac{\log C_n - \log C_0}{\log(1+i)}.$$

These formulae are used to calculate both the future value of a sum of capital invested at a given percentage and the annual interest that will have been earned by a final sum of capital based on an initial amount. They can also calculate the number of years or periods that must pass to reach a specific final sum of capital given an initial amount.

annual equivalent rate:

$$\text{Annual equivalent rate} = (\text{AER}) = \left[\left(1+\frac{i/100}{n}\right)^{n} - 1\right]\cdot 100.$$

If $i = 12\%$ per year, but is applied in 12 monthly payments ($n = 12$), the annual equivalent interest rate will be:

$$\text{Annual equivalent rate} = (\text{AER}) = \left[\left(1+\frac{12/100}{12}\right)^{12} - 1\right]\cdot 100 = 12.682503\%,$$

where $i = 12\%$ annual, $n = 12$ months.

If the interest rate was applied quarterly, it would be:

$$\text{Annual equivalent rate} = (\text{AER}) = \left[\left(1+\frac{12/100}{4}\right)^{4} - 1\right]\cdot 100 = 12.550881\%,$$

where $i = 12\%$ annual, $n = 4$ quarters.

The nominal interest rate is affected by the effects of inflation. For example, if an investor has invested in public debt bonds that pay an interest rate of 5%, and inflation is 3%, the real interest rate, that is to say the rate that determines the increase in real purchasing power provided by the return on their certificates, will be given, generally speaking, by the difference between the nominal interest rate and the rate of inflation.

$$\text{Real interest rate} = \text{Nominal interest rate} - \text{rate of inflation}.$$

How long will it take to double the invested capital?

Applying the formula for compound interest is simple. It allows us to determine the end value of an investment at a given interest rate over a specific period. For example, if an initial sum of capital, $C_0 = £10{,}000$ is invested for three years at an annual interest rate of 5%, what will be the final value of the capital, C_3?

$$C_0 = £10{,}000;\ i = 5\%\ (0.05),\ n = 3\ \text{years}.$$

Applying the formula $C_3 = C_0 \cdot (1 + i)^3$ gives:

$$C_3 = 10{,}000 \cdot (1 + 0.05)^3 = 10{,}000 \cdot 1.157625 = £11{,}576.25.$$

However, the problems created by the application of compound interest become more complicated when other variables of the problem become unknowns that must be solved. Hence, for example, one of the problems that may arise for an investor is knowing how long they will need to wait to double the value of a sum of capital invested at a specific interest rate, or reach a certain sum of money.

Consider this simple example: we wish to calculate how long it will take for the value of a sum of £10,000 of capital to double, applying a fixed interest rate of $i = 5\%$. Given the initial value of capital $C_0 = £10{,}000$ and the final value $C_n = £20{,}000$, and the interest rate $i = 5\%$, we can apply the following formula:

$$n = \frac{\log C_n - \log C_0}{\log(1+i)},$$

which gives us the following result:

$$n = \frac{\log 20{,}000 - \log 10{,}000}{\log(1+0.05)} = \frac{4.30103 - 4.00}{0.0211892991} = 14 \text{ years}$$

HOW MUCH INTEREST AM I BEING CHARGED?

The buyer of a car may ask what interest rate they are really being charged on their hire-purchase scheme. The salesperson tells them the price of the car is £10,000, which must be paid over five years, such that the total value for the amortisation of the capital and interest will be £15,000. The buyer wishes to calculate the type of interest applied.

Given the number of years $n = 5$, the initial value $C_0 = £10{,}000$ and the final value $C_n = £15{,}000$, it is possible to calculate the interest rate i using the following formula:

$$i = \sqrt[n]{\frac{C_n}{C_0}} - 1.$$

Applying the values of the problem in question, the interest rate that is applied will be:

$$i = \sqrt[5]{\frac{15{,}000}{10{,}000}} - 1 = \sqrt[5]{1.5} - 1 = 1.08447 - 1 = 0.08447 = 8.447\%.$$

Logarithms can be processed on a scientific calculator, or using mathematical formulae in a spreadsheet, or directly using a search engine (search log x).

How to calculate the instalments on a loan

Normally, if consumers or businesses do not have enough money to purchase long-lasting consumer goods or machinery and industrial or business facilities, they request a loan from a bank. For the purpose of property acquisition, the loans provided are guaranteed on the property that is acquired, or rather, they are assets with a mortgage guarantee or mortgaged assets. This means that if the debtor is unable to meet the obligations of the loan, the ownership of the property reverts to the bank that has provided the money. The repayment of mortgaged credit and loans is referred to as amortisation and involves a periodical payment (e.g. monthly, quarterly, twice yearly, annually) to the lender in an instalment that comprises two parts: an interest payment and the repayment of the capital that is borrowed.

The majority of mortgage credit and loans are repaid in constant instalments, whereby identical payments are made periodically and include variable parts to cover the interest and amortisation of the capital that has been loaned. The instalments can be paid at the beginning or end of each period, meaning that both the value of interest accrued and the amortisation of the capital will vary. Normally, the amortisation instalments of a loan are paid at the end of each period.

However, there are other systems for the amortisation of loans. There may be grace periods in which only interest is paid as variable instalments, in which a fixed part of the capital is amortised in each period in addition to the interest for the capital pending amortisation. In this system (the German system or constant amortisation) the periodical instalments are variable, since they include the fixed part for amortisation of the capital and the variable part for interest.

However, the most common loan amortisation system in use is the French system, or progressive amortisation. It entails specifying a constant instalment (normally payable at the end of each period), in which one part corresponds to the payment of interest and the other to the amortisation of the capital. In the first years for paying off the loan, the majority of the instalment comprises the payment of interest, and only a small part corresponds to amortisation of the capital. However, as time passes, since there is less capital to be returned, the part corresponding to the payment of interest decreases and the part corresponding to amortisation of the capital increases. Now let us consider how to calculate the value of the constant annuity for the amortisation of a loan of C_0 at an interest rate i over n periods (years). To carry out this calculation, we need to use the formula for the sum of the terms of a geometric progression.

A geometric progression is a series of numbers in which each term is given by the previous term multiplied by another fixed number referred to as the ratio of the series. Hence, the series of numbers $a_1, a_2, a_3, a_4 \ldots a_{n-1}, a_n$ (in which the subscript index indicates the term's position in the order: 1 for the first, n for the last) is a geometric progression, provided that, given a ratio r, the following holds: $a_2 = a_1 \cdot r$, $a_3 = a_2 \cdot r \ldots a_n = a_{n-1} \cdot r$, such that $r = a_n / a_{n-1}$. As a result, expressing the series for the geometric progression as a function of the first element, a_1, gives:

$$\begin{aligned}
&a_1 = a_1 \\
&a_2 = a_1 \cdot r \\
&a_3 = a_1 \cdot r^2 \\
&\ldots \\
&a_{\mathrm{n}} = a_1 \cdot r^{n-1}.
\end{aligned}$$

Defining S_n as the sum of this geometric progression, gives:

$$S_n = a_1 + a_2 + a_3 + \ldots + a_{n-1} + a_n. \tag{1}$$

Multiplying the two terms of equation (1) by the ratio r gives:

$$r \cdot S_n = r \cdot (a_1 + a_2 + a_3 + \ldots + a_{n-1} + a_n) = r \cdot a_1 + r \cdot a_2 + r \cdot a_3 + \ldots + r \cdot a_{n-1} + r \cdot a_n.$$

$$r \cdot S_n = a_2 + a_3 + \ldots + a_n + r \cdot a_n. \tag{2}$$

(since multiplying a term a_i by the ratio r gives the next term a_{i+1}, hence, $a_{i+1} = r \cdot a_i$).

Subtracting $r \cdot S_n - S_n$, or rather (2) – (1), gives:

$$r \cdot S_n - S_n = -a_1 + r \cdot a_n;\ S_n \cdot (r - 1) = r \cdot a_n - a_1,$$

where

$$S_n = \frac{r \cdot a_n - a_1}{r - 1}. \tag{3}$$

which is the formula for the summation of a geometric progression. Keeping in mind that $a_n = a_1 \cdot r^{n-1}$ and substituting this value in (3) gives:

$$S_n = \frac{r \cdot a_1 \cdot r^{n-1} - a_1}{r-1} = \frac{a_1 \cdot r^n - a_1}{r-1} = a_1 \cdot \frac{r^n - 1}{r-1};$$

which is another way of expressing the formula for the summation of the geometric progression:

$$S_n = a_1 \cdot \frac{r^n - 1}{r-1}. \tag{4}$$

For a loan with an annuity of a that remains constant for a period n with an interest rate i, the future value C_n of the payment of the n annuities of a will be:

$$\begin{aligned} C_n &= a \cdot (1+i)^0 + a \cdot (1+i)^1 + \ldots + a \cdot (1+i)^{n-2} + a \cdot (1+i)^{n-1} = \\ &= a + a \cdot (1+i)^1 + \ldots + a \cdot (1+i)^{n-2} + a \cdot (1+i)^{n-1}. \end{aligned}$$

The result is the summation of a geometric progression with a ratio of $(1+i)$ and whose first term is a.

Applying the formula $S_n = a_1 \cdot \frac{r^n - 1}{r-1}$ (4) for the sum of geometric progression gives:

$$C_n = a \cdot \frac{(1+i)^n - 1}{1+i-1} = a \cdot \frac{(1+i)^n - 1}{i}. \tag{5}$$

Since $C_n = C_0(1+i)^n$, substituting this value in (5) gives:

$$C_0(1+i)^n = a \cdot \frac{(1+i)^n - 1}{i}.$$

And solving for a, which is the constant annuity, gives the formula for calculating the constant annuity a of a loan:

$$a = \frac{C_0 \cdot i \cdot (1+i)^n}{(1+i)^n - 1}, \tag{6}$$

where C_0 is the value of the loan.

SEQUENCES

The geometric progression is one of the simplest types of sequences. A sequence is an ordered set of numbers in which it is possible to determine the value of a given term using a mathematical formula that contains a variable indicating the position occupied by the term in the sequence. The formula is called the general term. This is a function of the position variables (often referred to as n). The formula for the general term is a function: $a_n = f(n)$, that depends on the position of the term n.

There is another type of sequence in which the value of a term is obtained using a formula that contains a number of previous terms. This is the case with the Fibonacci sequence, 1, 1, 2, 3, 5, 8, 13..., in which each term is the sum of the two previous terms. Or the sequence of the general term $a_n = n + a_{n-1}$; $a_1 = 3$, which is 3, 5, 8, 12, 17, 23...

For all sequences, it is necessary to indicate the value of the initial term or terms, as well as the number of terms, if it is a limited sequence.

A sequence can consist of a limited or infinite number of terms, and in the latter case, we can extend the sequence for as long as we wish, using the formula for the general term to obtain new terms. There are increasing sequences (where the value of the terms increases) and decreasing sequences (where the value decreases), and sequences can be limited and unlimited.

They are widely used in financial mathematics. Hence, the sequence that gives the interest accrued by an initial unit of capital with a simple interest rate of 20% per year, after one year, two years, etc. is 1; 1.2; 1.4; 1.6; 1.8; 2.0; 2.2... This is an increasing sequence, the general term of which is $a_n = 1 + 0.2 \cdot n$.

The sequence that provides the interest accrued by an initial unit of capital with a compound interest rate of 20% per year, after one year, two years, etc. is 1; 1.2^2; 1.2^3; 1.2^4... This is an unlimited increasing sequence whose general term is $a_n = (1 + 0,2)^n$.

The sequence 21, 23, 25, 27, 29, 31... is an unlimited increasing sequence whose general term is $a_n = 21 + 2\,(n - 1)$; $a_1 = 21$.

The sequence 1, 5, 25, 125, 625, 3,125... is an unlimited increasing sequence whose general term is $a_n = 5^{n-1}$; $a_1 = 1$.

The sequence 1, $\frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \frac{1}{9}...$ is an unlimited decreasing sequence whose general term is $a_n = \frac{1}{2n-1}$; $a_1 = 1$.

Finally, 1, $\frac{1}{7}, \frac{1}{49}, \frac{1}{343}, \frac{1}{2.401}...$ is an unlimited decreasing sequence whose general term is $a_n = \frac{1}{7^{n-1}}$; $a_1 = 1$.

How to reduce the instalments on debts

When a loan is taken out, a contract is signed with the lender establishing the conditions of the credit, such as instalments, frequency of payments, applicable interest rate, annual equivalent rate (for payments of less than one year), in addition to the conditions established in the event that one of the parties fails to comply.

When payments are to be made at the end of each period, the fixed instalment will be calculated using the formula deduced in the previous section. The fixed instalment comprises one part for the payment of the interest accrued during the period in question and another for the amortisation of the capital. At the end of each period, the remaining capital will decrease, meaning that the value of the interest will also decrease, and conversely, the proportion destined for the amortisation of the capital will progressively increase over subsequent periods.

This information can be used to calculate the amortisation table for a loan, making it possible to know the amortised capital and that remaining at any point throughout the lifespan of the loan. As an example, the amortisation table of a £10,000 loan at an annual rate of 5% for 5 years is provided below. The fixed annual instalment corresponding to these terms of credit can be calculated to give a value of £2,309.75 per year.

Applying formula (6) gives the fixed instalment of the loan:

$$a = C_0 \cdot \frac{i \cdot (1+i)^n}{(1+i)^n - 1} = 10,000 \cdot \frac{0,05 \cdot (1.05)^5}{(1.05)^5 - 1} = 10,000 \cdot \frac{0.06381408}{0.27628156};$$

$$a = 10,000 \cdot 0.2309748 = 2,309.75.$$

Year	Capital	Interest	Sum of Interest	Capital Amortisation	Fixed Instalment	Outstanding Capital
1	10,000.00	0.05	500.00	1,809.75	2,309.75	8,190.25
2	8,190.25	0.05	409.51	1,900.24	2,309.75	6,290.01
3	6,290.01	0.05	314.50	1,995.25	2,309.75	4,294.76
4	4,294.76	0.05	214.74	2,095.01	2,309.75	2,199.75
5	2,199.75	0.05	109.99	2,199.76	2,309.75	–0.01

Amortisation table for loan 1.

As can be seen, as the years pass and the instalments are paid, the outstanding capital decreases, and the value of the interest also decreases as the amount designated for amortising the capital increases.

However, it may also be the case that an individual or family has a number of outstanding loans, and it becomes extremely difficult for them to meet the payments with their level of income. Hence, for example, if the person who has taken out the previous loan has another for £30,000 over 10 years at an annual rate of 8%, for which they pay an instalment of £4,470.88, the total value of the instalments they are paying will be £6,780.63.

Year	Capital	Interest	Sum of Interest	Capital Amortisation	Fixed Instalment	Outstanding Capital
1	30,000.00	0.08	2,400.00	2,070.88	4,470.88	27,929.12
2	27,929.12	0.08	2,234.33	2,236.55	4,470.88	25,692.57
3	25,692.57	0.08	2,055.41	2,415.47	4,470.88	23,277.10
4	23,277.10	0.08	1,862.17	2,608.71	4,470.88	20,668.38
5	20,668.38	0.08	1,653.47	2,817.41	4,470.88	17,850.97
6	17,850.97	0.08	1,428.08	3,042.80	4,470.88	14,808.17
7	14,808.17	0.08	1,184.65	3,286.23	4,470.88	11,521.94
8	11,521.94	0.08	921.76	3,549.12	4,470.88	7,972.82
9	7,972.82	0.08	637.83	3,833.05	4,470.88	4,139.77
10	4,139.77	0.08	331.18	4,139.70	4,470.88	0.07

Amortisation table for loan 2.

If the person finds it extremely difficult to meet these payments, they may find themselves required to ask the financial entity or another bank to refinance their loans at a lower interest rate, but above all with a lower total instalment, because, for example, they are unable to pay more than £5,000 a year The new financial entity may offer to consolidate the two loans into a single loan for £40,000 with an interest rate of 6%. The problem lies in determining the duration of the new loan, since the recipient is only able to pay an instalment of £5,000.

To calculate the duration of the new loan, it is necessary to solve formula (6) for calculating the amortisation instalment of a loan in terms of the variable *n*.

$$a = C_0 \cdot \frac{i \cdot (1+i)^n}{(1+i)^n - 1}.$$

Dividing both terms by C_0, gives

$$\frac{a}{C_0}=\frac{i\cdot(1+i)^n}{(1+i)^n-1}.$$

And dividing the two terms by $(1 + i)^n$ gives:

$$\frac{a}{C_0}=\frac{\frac{i\cdot(1+i)^n}{(1+i)^n}}{1-\frac{1}{(1+i)^n}};\ \frac{a}{C_0\cdot i}=\frac{1}{1-\frac{1}{(1+i)^n}};\ 1-\frac{1}{(1+i)^n}=\frac{C_0\cdot i}{a};\ 1\text{-}\frac{C_0\cdot i}{a}=\frac{1}{(1+i)^n};$$

$$\frac{1}{(1+i)^n}=\frac{a-C_0\cdot i}{a};\ (1+i)^n=\frac{a}{a-C_0\cdot i}.$$

We can now make use of logarithms:

$$\log(1+i)^n=\log\left(\frac{a}{a-C_0\cdot i}\right);$$

$$n\cdot\log(1+i)=\log(a)-\log(a-C_0\cdot i).$$

to solve for the number of years, *n*:

$$n=\frac{\log(a)-\log(a-C_0\cdot i)}{\log(1+i)}.$$

Finally, applying this formula and substituting *a*, *i* and C_0 for the corresponding values gives:

$$n=\frac{\log(5{,}000)-\log(5{,}000-40{,}000\cdot 0.06)}{\log(1.06)}=\frac{3.69897-3.41497335}{0.02530587}=11.22\text{ years.}$$

Drawing up the amortisation table, it can be seen that £1,073.73 of the capital still needs to be amortised in year 11, which represents 21.5% of the instalment, the same as saying that the duration of the loan is 11 years and 22% of a year, in other words 11 years and 80 days.

Year	Capital	Interest	Sum of Interest	Capital Amortisation	Fixed Instalment	Outstanding Capital
1	40,000.00	0.06	2,400.00	2,600.00	5,000.00	37,400.00
2	37,400.00	0.06	2,244.00	2,756.00	5,000.00	34,644.00
3	34,644.00	0.06	2,078.64	2,921.36	5,000.00	31,722.64
4	31,722.64	0.06	1,903.36	3,096.64	5,000.00	28,626.00
5	28,626.00	0.06	1,717.56	3,282.44	5,000.00	25,343.56
6	25,343.56	0.06	1,520.61	3,479.39	5,000.00	21,864.17
7	21,864.17	0.06	1,311.85	3,688.15	5,000.00	18,176.02
8	18,176.02	0.06	1,090.56	3,909.44	5,000.00	14,266.58
9	14,266.58	0.06	856.00	4,144.00	5,000.00	10,122.58
10	10,122.58	0.06	607.35	4,392.65	5,000.00	5,729.93
11	5,729.93	0.06	343.80	4,656.20	5,000.00	1,073.73

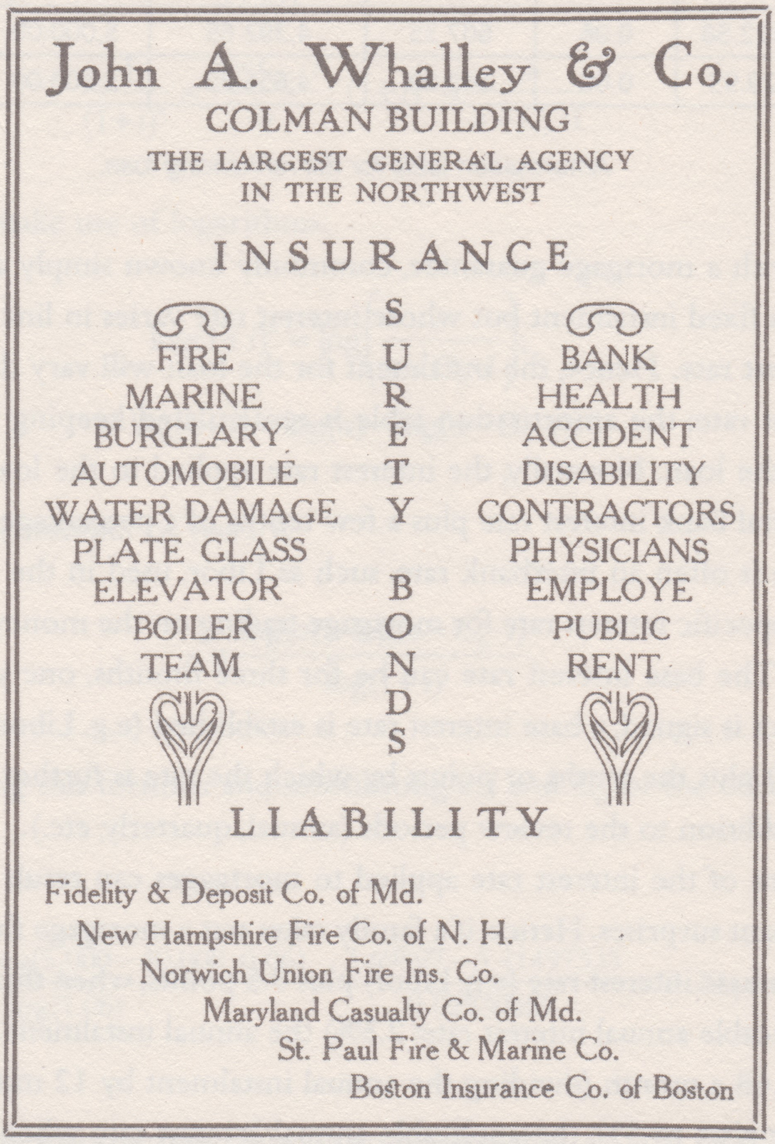
Amortisation table for the refinancing loan.

Loans with a mortgage guarantee, commonly known simply as mortgages, are loans with a fixed instalment but whose interest rate varies in line with changes in a base interest rate. Hence, the instalment for the loan will vary depending on the new interest rate; the amortisation table is recalculated keeping the same expiry period for the loan. Normally, the interest rate applied to the loan is based on an official central bank interest rate plus a few tenths of a percentage point. This total interest rate is often an interbank rate, such as Libor, used in the City of London, although a specific interest rate for mortgage trading on the money market can also be applied. The base interest rate can be for three months, one year, etc. When a loan contract is signed, a base interest rate is established (e.g. Libor at three months or one year) plus the tenths or points by which the rate is further increased by the lender, in addition to the review periods (annual, quarterly, etc.).

Revisions of the interest rate applied to mortgages can result in both pleasant and unpleasant surprises. Hence, if a family takes out a mortgage for £300,000 over 20 years at a base interest rate (e.g. Libor) plus 0.5 points, when this interest rate was at 2% (applicable annual interest rate: 2.5%) the annual instalment was £19,244.14, i.e. £1,603.68 a month (dividing the annual instalment by 12 months). If the base interest rate rises to 5%, the applicable annual interest rate when the loan is next reviewed will be 5.5%, causing the instalment to increase to £25,103.80 (which represents £2,091.98 per month), hence it has increased by 30.45%. Conversely, when the interest rate falls, the instalments decrease considerably.

Fundamentals of insurance

Insurance companies cover risks for businesses, families and individuals in many fields, such as accidents, car breakdown, illness and death (life insurance), burglary, medical cover and damages corresponding to numerous types of insurance risks (e.g. fires, floods, climate and civil liability). In order to cover these risks, insurance companies calculate the prices they must charge their customers, such that, taking into account the provisions for the necessary expenditure to cover indemnifying any accidents that occur, the insurance activity remain profitable.



John A. Whalley & Co.
COLMAN BUILDING
THE LARGEST GENERAL AGENCY
IN THE NORTHWEST

INSURANCE

FIRE
MARINE
BURGLARY
AUTOMOBILE
WATER DAMAGE
PLATE GLASS
ELEVATOR
BOILER
TEAM

SURETY BONDS

BANK
HEALTH
ACCIDENT
DISABILITY
CONTRACTORS
PHYSICIANS
EMPLOYE
PUBLIC
RENT

LIABILITY

Fidelity & Deposit Co. of Md.
New Hampshire Fire Co. of N. H.
Norwich Union Fire Ins. Co.
Maryland Casualty Co. of Md.
St. Paul Fire & Marine Co.
Boston Insurance Co. of Boston

Advert for an American insurance company at the start of the 20th century advertising the services it offered the public.

Premiums are calculated based on statistical methods, such as sampling, and mathematical methods, such as actuarial calculations, curve fitting (linear and non-linear regression), probability theory, and matrix calculations for mortality tables and population death rates. The application of these methods must be grounded in the preparation of reliable historical data, since in order to predict future events it is necessary to have extremely rigorous information about past changes in the different variables that affect the future determination of costs originating from the coverage of the insured accidents. These can then be used to establish the premium to be paid by the customer.

Insurance companies are interested in studying whole populations to make their forecasts for accident rates. However, the large size of certain populations requires them to make use of statistical sampling techniques, or rather focus on a small part of the population that is being studied, whose properties are similar to the larger population, meaning that the level of confidence in the sample must be determined, in addition to the sampling error. There is a wide range of sampling techniques available, which can be based on random criteria or stratified sampling, whereby representative samples are selected from various segments of the population (or sectors), different geographic areas, etc.

Insurers also prepare statistics on the cost of accidents, the total number of accidents, the total number of policies and revenue that is generated, making comparisons between the payments made for accidents and the revenue derived from premiums. These basic variables are combined with others that also affect the insurance, such as age or geographic area, in order to determine formulae that make it possible to predict accident rates, risks, etc. through the application of multivariable, cluster and chaid (decision tree) analysis techniques.

The actuarial calculation of life insurance premiums, is based on tables (matrices) for the probability of death and survival, in which the probability of reaching the next group alive is determined in historical series for age group and sex. Given the costs of coverage for the accidents, it is possible to determine the value of the premiums in order to ensure the profitability of the business activity of insuring them.

Descriptive statistics

The science of statistics makes it possible to study situations, things or people by means of quantification. The perspective of such situations, things or people is thus no longer exclusively intuitive and acquires a certain objectivity. Naturally, it is possible to play many tricks with statistics so they can be deceptive. However, a well-planned and executed statistical study results in a more accurate understanding of reality and facilitates better decision-making to influence it.

Statistics can be defined as the operation of mathematical analysis that makes it possible to study phenomena for which incomplete knowledge is available with maximum precision. The economy is doubtless an excellent example of such a phenomenon. Three key concepts are characteristic of the said phenomena to be studied – specifically to analyse its variability: the individual; each element affected in a measurement, a number in an order or an index of the property being studied (e.g. people, families, voters, articles, units or dust mites); and the population, the set of individuals for which we wish to undertake a statistical study.

When a population is analysed, values for the characteristic for all its individuals can be considered. However, in many cases this will be impractical, making it necessary to select a group of individuals from the population, referred to as the sample, and which is made up of the individuals that are really observed. In fact, one of the problems that must be solved using statistics is the selection of the sample and how to study it in order to draw valid conclusions for the population as a whole.

The properties that are studied can be qualitative (e.g. the level of satisfaction with respect to a given service) or quantitative. They can be discrete (i.e. can only have integer values, such as the number of children per family) or continuous (i.e. the diameter of the parts manufactured by different machines). The data recorded in the statistical study of a variable is displayed using graphs, of which there are many different types: bar charts, histograms, line graphs, pie charts, etc.

Chapter 4

The Costs of Production and Returns on Investment

The maximum level of production that can be obtained from a given quantity of factors is defined by the following function: $x = f(v_1, v_2 \ldots v_n)$, where x is the quantity produced and $v_1, v_2 \ldots v_n$ are the quantities of the factors of production. These quantities are often determined by the level of technology and in principle do not vary (assuming the production technique remains the same) until there is a change in technology, making it necessary to change the quantities of the factors used, or replace them with more economical and efficient alternatives. The production function implies that the variables it includes (quantities of the product and the factors of production) are measurable, and also that the production technique remains the same and the behaviour of the factors is continuous.

The optimal production level

Like all functions, the production function can be expressed analytically by a formula or graph, or numerically using a statistical table. The latter entails arranging the quantities of the factors of production required for a given quantity of the product into rows and columns. The graph represents the values contained in the table.

To be able to set out a simplified model with just two factors of production, let us now consider an example based on the manufacture of an aluminium and nickel alloy in which the final level of production (x) is a function of the quantity of the nickel (v_1) and the aluminium (v_2), expressed by the following formula:

$$x = v_1 + 2v_2.$$

FUNCTIONS

A function is a quantitative relationship between variables. In the simplest case, it is a matter of determining the influence of various values of a variable (referred to as independent, because its values are selected at will) on the corresponding values of the other variable (referred to as dependent since it is the result of the values given to the first). There are empirical functions, where data comes from experiments, and mathematical functions with data governed by an exact formula.

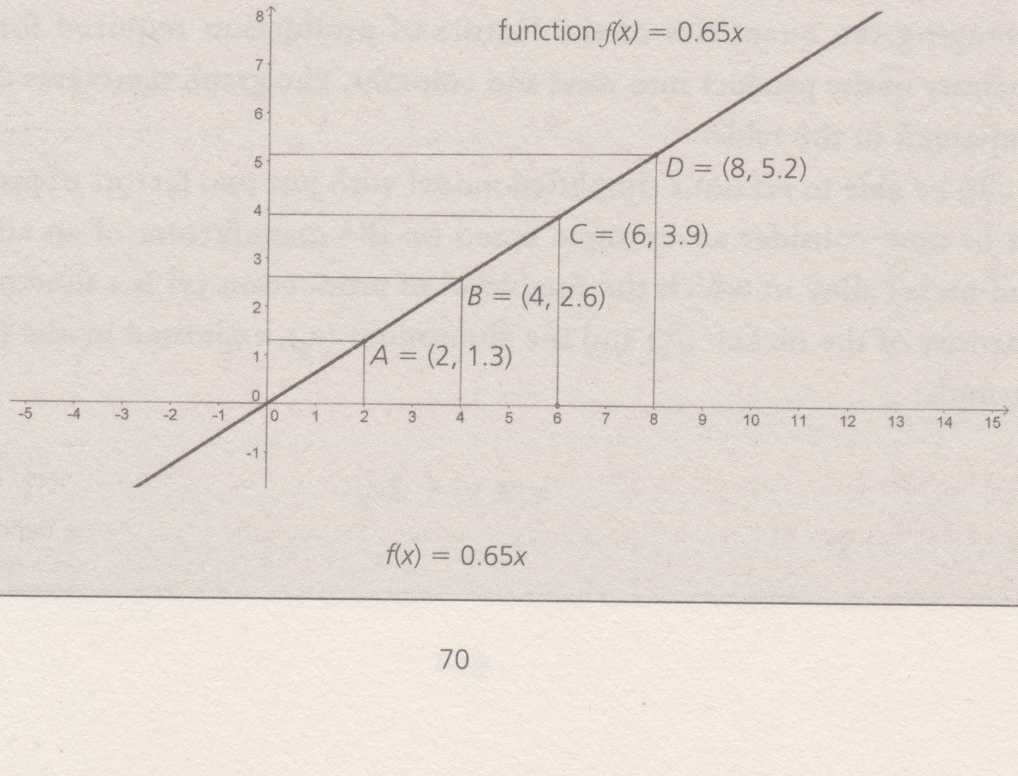
In economics, certain empirical functions that relate variables can be approximated by a mathematical function using a technique called regression. On other occasions, there is a formula that directly provides a suitable approximation of the values corresponding to the two variables. If we consider the price to be paid as the dependent variable and the number of units purchased as the independent variable, the mathematical function represents a relationship that can be expressed as follows:

$$\text{Quantity to be paid} = \text{unit price} \cdot \text{units acquired.}$$

In mathematical notation, this is represented as $f(x) = a \cdot x$, where $f(x)$ indicates that the value of f depends on x (dependent variable), a is a constant value representing the unit price and x is the number of units acquired (independent variable).

There are three ways to 'describe' a mathematical function: using its formula, using a table of values, or by drawing its graph on a system of Cartesian coordinates. There are many different types of functions. The simplest are linear functions, or first-degree polynomials, such as $f(x) = 0.65x$. The three forms of representing the function are shown below.

x	$f(x)$
0	0
2	1.3
4	2.6
6	3.9
8	5.2

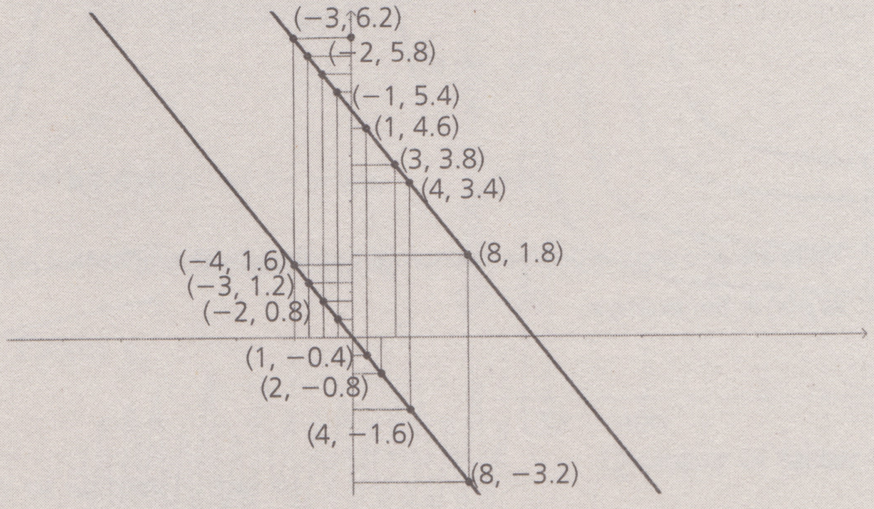


$f(x) = 0.65x$

The function $f(x) = 0.65x$ is an increasing linear function, or one that is directly proportional, and represents the simple direct rule of three. In this function, increasing the independent variable x increases the dependent variable $f(x)$. When x jumps from 2 to 4, the function $f(x)$ jumps from 1.3 to 2.6. For the practical function we proposed above (quantity to be paid = unit price · units purchased), negative values of units purchased do not make sense and the part of the graph located below zero is meaningless.

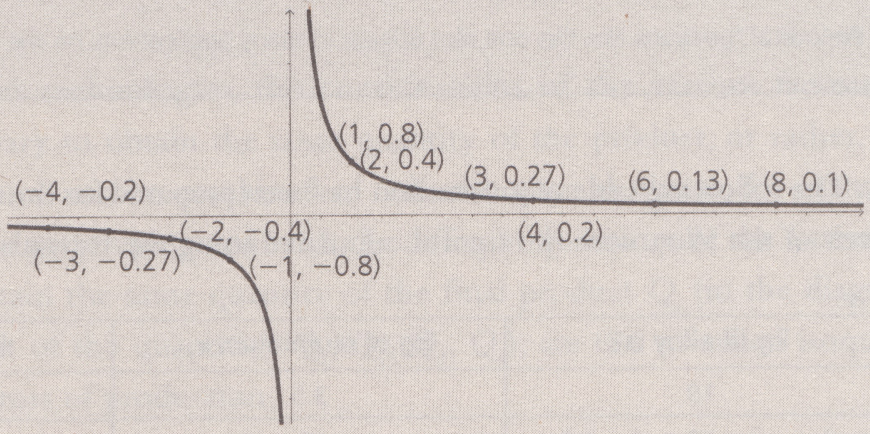
There are many other linear functions. Each indicates a different relationship between the two variables x and $f(x)$, such as the two decreasing linear functions $g(x)$ and $h(x)$ in the following examples.

x	$g(x)$	$h(x)$
−4	1.6	6.6
−3	1.2	6.2
−2	0.8	5.8
−1	0.4	5.4
0	0	5
1	−0.4	4.6
2	−0.8	4.2
3	−1.2	3.8
4	−1.6	3.4
5	−2	3
6	−2.4	2.6
7	−2.8	2.2
8	−3.2	1.8



$g(x) = -0.4x \qquad h(x) = -0.4x + 5$

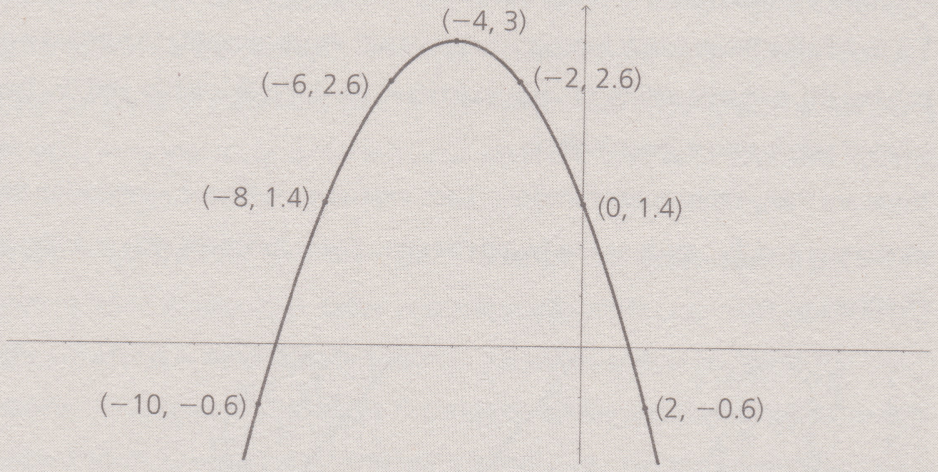
x	$i(x)$
−4	−0.2
−3	−0.267
−2	−0.4
−1	−0.8
0	---
1	0.8
2	0.4
3	0.267
4	0.2
6	0.133
8	0.1



$i(x) = 0.8/x$

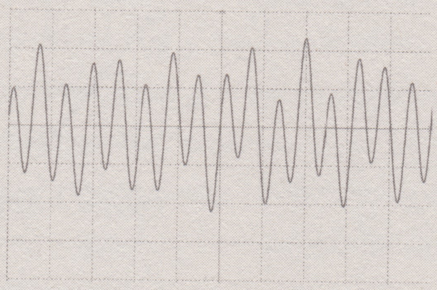
The function $i(x)$ is the inverse proportionality function. It represents the simple inverse rule of three. For this type of function, increasing the independent variable x, causes the dependent variable $i(x)$ to decrease in an inversely proportional manner. For example, when x is 3, $i(x)$ is $0.8/3 = 0.267$.

x	$p(x)$
−10	−0.60
−8	1.40
−6	2.60
−4	3.00
−2	2.60
0	1.40
2	−0.60



$$p(x) = -0.1x^2 - 0.8x + 1.4$$

There are also second-degree polynomial functions, with graphs that are parabolic curves, such as $p(x)$ in the example.



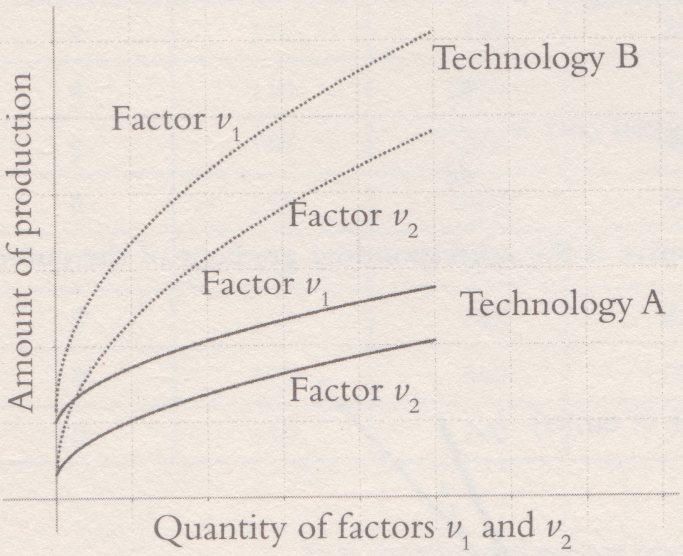
Periodical functions like this one also appear in many applications on the stock market.

The following table of production has been prepared for the different production levels of the alloy, with the figures calculated using the formula

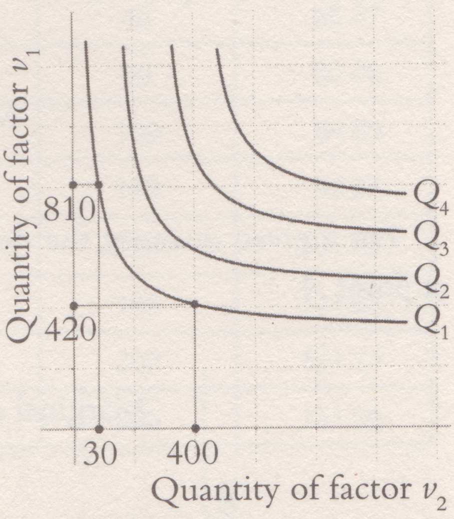
kg of alloy (x)	kg of aluminium v_2	kg of nickel v_1
10	3.5	3
20	7	6
30	10.5	9
40	14	12
50	17.5	15

Combination of the factors of production for the manufacture of an aluminium-nickel alloy.

Any change in technology will result in modifications to the combinations of the two production factors and hence lead to a new table of production with the corresponding change in the production function. For each technology scenario, there is a production graph with its own factor curves that relate the quantity of the final product to the quantities of each factor for this technology scenario. Hence, for example, in the graph shown below, the left graph shows the changes in the combinations of the factors of production v_1 and v_2 corresponding to the two different technology scenarios, *A* and *B*.



Production curves.



Isoquant curves.

Using different technologies, the combinations of the various factors of production will vary to obtain the same quantity of the product, or rather, the factors of production can be combined in different ways to obtain, for example, 50 kg of alloy. The graphs prepared with the different quantities of the factors of production to obtain the same quantity of the final product Q_i (in the diagram above to the right of the quantities Q_1, Q_2, Q_3, Q_4), are the so-called isoquant curves for the factors of production.

Each isoquant curve represents the various quantities of the factors of production that generate the same final quantity Q_i of production. For example, the quantity of production Q_1 can be obtained by the combination of 420 units of factor v_1 and 400 units of factor v_2, or by the combination of 810 units of factor v_1 and 30 units of factor v_2.

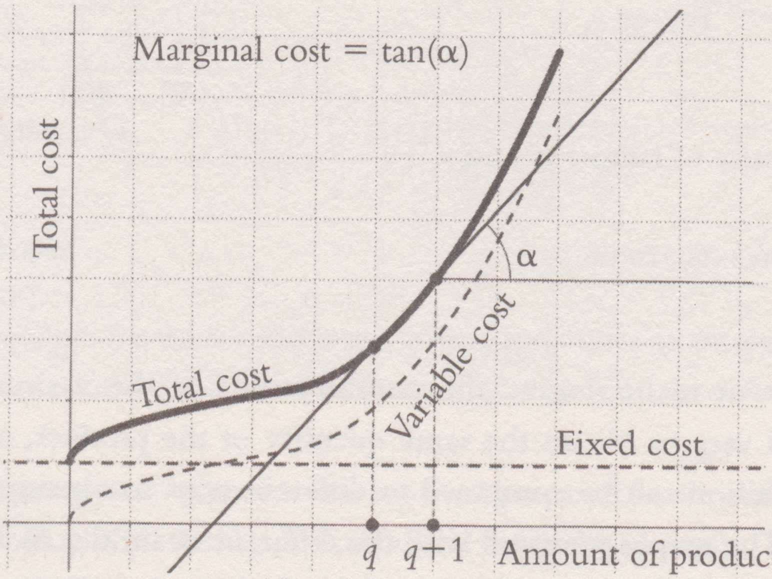
Any production process entails fixed costs and variable costs that depend on the quantity produced, with the sum of these two quantities representing the final cost. The marginal cost is defined as the total additional cost generated by increasing the production by an extra unit of the final product.

$$\text{Marginal cost} = \frac{\Delta \text{ costs}}{\Delta \text{ of producing one unit}}.$$

It is also expressed as the derivative of the cost function with respect to the derivative of the quantity of the product:

$$\text{Marginal cost} = \frac{dc}{dq}.$$

For a given quantity, the derivative is the corresponding gradient of the curve at point *A*.



Curve for total cost with the derivative and gradient of the curve.

The average or unit cost is the result of dividing the total cost by the number of units produced:

$$\text{Average cost} = \frac{\text{Total cost}}{\text{Total production}}.$$

For example, in a production run of 10 units of product, the marginal cost and average cost of each unit are calculated as follows:

Quantity	Fixed Cost	Variable Cost	Total Cost	Marginal Cost	Average Cost
0	110	0	110		
1	110	60	170	60	170.00
2	110	110	220	50	110.00
3	110	150	260	40	86.67
4	110	210	320	60	80.00
5	110	310	420	100	84.00
6	110	450	560	140	93.33
7	110	600	710	150	101.43
8	110	780	890	180	111.25
9	110	980	1,090	200	121.11
10	110	1,200	1,310	220	131.00

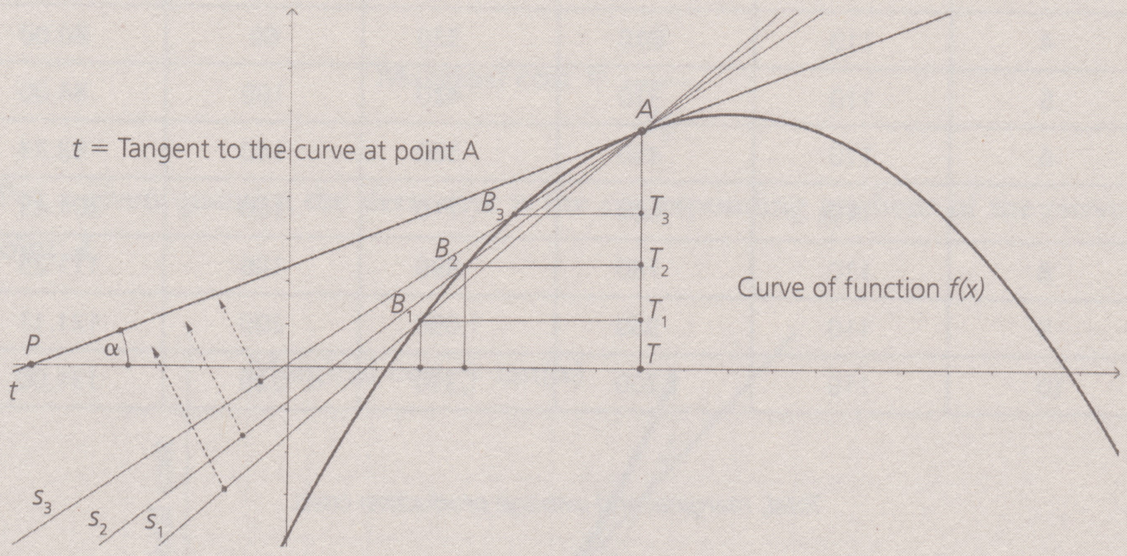
Total, marginal and average production costs.

The information allows us to draw the curves for the production, as well as marginal and average costs. As we can see, the average cost is often much higher than the marginal cost. The curve for the marginal cost indicates the minimum costs for each quantity of product. The point at which the marginal cost and average cost curves intersect coincides with the minimum point of the average cost. While the marginal cost is lower than the average cost, it is decreasing, whereas from the point at which they intersect, the marginal cost is greater than the average cost, with the latter increasing.

The tendency of the average cost curve to rise from the point at which it meets the marginal cost curve is responsible for the law of diminishing returns. This is when the average cost curve decreases until it crosses the marginal cost curve, the point that determines its minimum, from which point onwards average costs will grow (provided changes in technology do not occur to make the cost of production cheaper) and, as such, as average costs increase returns will diminish.

TRENDS AND DERIVATIVES

In economics, to determine the trend of one variable with respect to another (such as the level of supply of a good with respect to its price), we must know how much the function that represents it is growing at a certain point *A* on its graph. This allows us to compare its level of variation (increasing or decreasing) with respect to the value at other points on the curve. We use the gradient of the point to express this level of variation. Therefore, we calculate the slope of the straight line tangent to the curve at point *A*. This is determined by calculating the angle at which the straight line tangent at *A* meets the horizontal axis. If the angle is extremely large, this means the straight line has a large gradient, and vice versa.

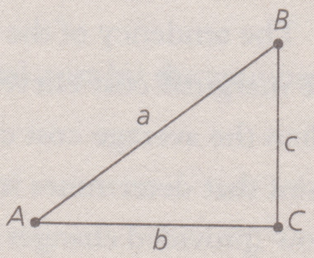


Let us consider various points on the curve: B_1, B_2, B_3, which grow increasingly closer to point *A*. Now observe the straight lines s_1, s_2, s_3, joining these points to *A* and intersecting the curve (secant lines). Note that the secants s_1, s_2, s_3 gradually approach *t*, the straight line tangent to the curve at point *A*.

The angles of the secant lines s_1, s_2, s_3 with the horizontal axis decrease until the points B_1, B_2, B_3 coincide with *A*, the angle with the horizontal axis is the same as the tangent line (α, alpha).

For a right-angled triangle *ABC*, the tangent of the angle *A* is $\tan(A)=\frac{c}{b}$.

The quotient of the two catheti for each triangle AB_1T_1, AB_2T_2, AB_3T_3 varies as the position changes and is the tangent of angle 1, angle 2, angle 3..., until the *B* points on their 'journey' towards *A* reach this point (the limit) and become the tangent of the angle (α, alpha) we have indicated.



$$\frac{AT_1}{B_1T_1} = \tan(\text{angle } 1);\ \frac{AT_2}{B_2T_2} = \tan(\text{angle } 2);\ \frac{AT_3}{B_3T_3} = \tan(\text{angle } 3);\ \ldots\ \frac{AT}{PT} = \tan\alpha.$$

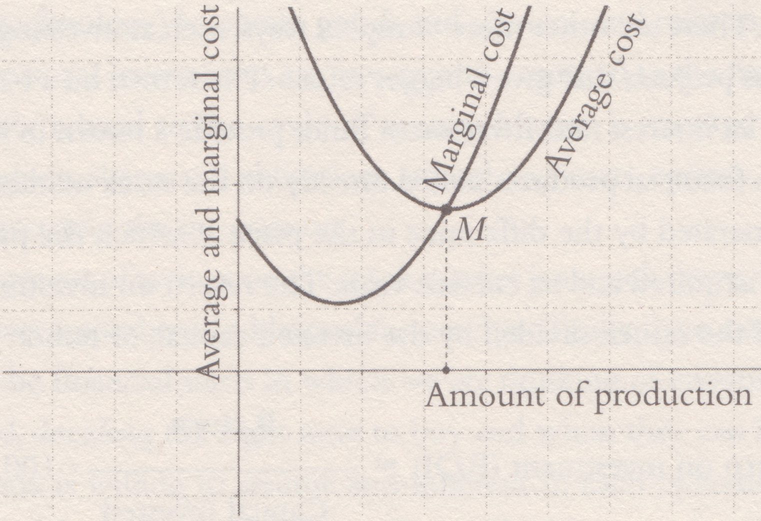
The $\lim\limits_{Bi \to A} \tan(\text{angle } i) = \frac{AT}{PT} = \tan\alpha$ is the gradient of the line t (tangent line of the function $f(x)$ at point A).

The value of $\lim\limits_{Bi \to A} \tan(\text{angle } i) = \frac{AT}{PT}$ is referred to as the derivative of the function $f(x)$ at point A and coincides with the value of the gradient of the tangent of the function at A.

The derivative of the function is calculated by applying the rules for derivation deduced from calculating the aforementioned limits. The tables of derivative functions for each type of function are used to solve individual cases in the study of trends.

$$f'(x_0) = \lim_{Bi \to A} \frac{AT_i}{B_iT_i} = \tan\alpha = \text{Derivative of } f(x) \text{ in } A(x_0, y_0) = \left(\frac{df(x)}{dx}\right)(x_0).$$

The point at which the average and marginal cost curves meet represents the optimal production level of a company, since it constitutes the minimum of the average cost curve and, logically, companies always aim to produce goods at the minimum cost.



Average and marginal cost curves.

However, the problem is not so straightforward, since in order to reach the point of optimal production, we must bear in mind that different technological setups give various combinations of the factors of production for producing the same quantity of a product. Which combination of the factors of production is optimal for the company?

Simplifying, let us assume we have only two factors, A and B, leading us to define a new concept: the marginal product for a factor of production A or B. Putting it another way, the increase in production generated by one unit of the factor of production (A or B) for each monetary unit of its cost.

When it comes to comparing the productivity of the two factors of production, the optimal production level occurs when the two marginal products are equal. Therefore:

$$\frac{\text{Marginal Product A}}{\text{Price of A}} = \frac{\text{Marginal Product B}}{\text{Price of B}}$$

To see a trend in a process (be it economic, sociological or statistical) for which the representative function *f*(*x*) is known, we can change the value of the variable *x* and take a look at the graph of the function which represents the process mathematically.

How and where to save: methods for evaluating investments

The easiest way to save money is through a fixed-term bank deposit or buying into investment funds. There are other more complex ways, such as investing in the stock market or business projects that give a bigger return. The return on a bank deposit is paid by means of an interest rate. Investment funds provide a return in the form of a dividend, just like financial products traded directly on the stock market, in addition to the surplus generated by the difference in the price at which the product trades on the day it was acquired and its current value. The return on investment (ROI) is the percentage of the return divided by the invested capital, or rather:

$$\text{Return on investment (ROI)} = \frac{\text{Return}}{\text{Capital invested}} \cdot 100.$$

The rate of return of the capital is used for measuring the performance of a financial security or a business or industrial project.

When an investment is made, it is also possible to compare the return obtained with what would have been obtained if we had invested the money in a bank account at a specific interest rate. Therefore we can determine if the difference between the returns of the two options suggests the investment is favourable. This return is referred to as residual income or added economic value and is based on making a profit B from a sum of capital that is invested C, and comparing the profit obtained B with what would have been obtained if the capital C had been invested in a bank account at an interest rate i, such that:

$$\text{Added economic value or residual income} = B - i \cdot C.$$

For example, if an investment of £1,000 provides a return of £110, and the interest rate offered by the bank is 5%, the return on investment and the added economic value will be:

$$\text{Return on investment} = \frac{110}{1{,}000} \cdot 100 = 11\%.$$

$$\text{Added economic value or residual income} = 110 - (1{,}000 \cdot 0.05) = £60.$$

When discussing investments, it is important to distinguish between real investments and financial investments. The former are made as part of commercial activities, combining the factors of production (e.g. capital, labour, raw materials, energy and technology) to obtain goods and services, and refer to the so-called real economy. Financial investments are investments confined to the financial system and are exclusively related to the cycle of money. Money is invested in bank deposits, financial products, shares and obligations, underlying assets, options, futures, derivative products, warrants... the list goes on.

The placement of savings on the stock market is a much more delicate operation, since it involves a number of additional factors. Specifically, these are knowledge of the market and the financial assets in which we are thinking of investing and which can be purchased; deciding the right time to buy and when they can be sold; awareness of the risk one is willing to assume and inherent in each operation; knowledge of and analysis of the evolution of the market with respect to the securities in which the investment will be made form part of the portfolio; and the financial solvency and future prospects of the assets, all of which will be considered in the next chapter.

In this chapter, we will focus our attention on real investments. Let us consider a simple example of a business, which with a sum of capital x (the unknown for this problem) makes an annual profit or constant income of a_i = £1,000. If the constant rate of return r on the capital x that is invested is known (e.g. $r = 5\%$), we must calculate the value of the capital.

Using the simple interest formula, we can deduce that:

$$a_i = r \cdot x,$$
$$x = a_i \,/\, r.$$

Hence, $1{,}000 = r \cdot x$, where $x = 1{,}000/0.05 = 20{,}000$. The sum of capital that will give a constant profit of £1,000 per year is £20,000.

All investments, both in business and industrial projects, involve an outlay of initial expenditure in the first year, in addition to making profits or losses for each year of the duration of the investment. Hence, it is a cash flow in which the value of the investment represents a negative flow (the expenditure) for the first year, and for the following years, when the business or project is up and running, positive flows (the profits), although in some years there may also be negative flows (losses) instead. All investments have a lifespan, just like living beings. Amortisations are calculated based on the duration of an investment (referred to as its useful life).

One of the problems that arises when making an investment is determining if sufficient cash flows will be generated during the useful life of the project to recover the investment that was made as a minimum. However, during the life of the project, it is clear that the purchasing power of money may be eroded by inflation, and hence the investor is not only interested in knowing if they will recover the money they have invested, but also if the present value of all forecasted cash flows, adjusted to take into account of the rate of inflation, will exceed the present value of the investment that is made.

It is clear that £100 today will be worth more than £100 in four years, and even more if we forecast an annual rate of inflation of 5% for the next four years. If, for example, we have an annual profit of £100 for today and the next four years, it is clear that the £100 for the coming year will be worth only 95% of what it is currently worth, since it will be affected by the rate of inflation. By the third year, the £100 that has been projected for the future will only be worth £90.25 at present value.

To summarise, a profit of £100 over five years with an annual rate of inflation of 5% (i.e. a starting total of £500), will have a present value of:

First year: 100.
Second year: 100/1.05 = 95.23809524.
Third year: 95.23809524/1.05 = 90.70294785.
Fourth year: 90.70294785/1.05 = 86.38375985.
Fifth year: 86.38375985/1.05 = 82.27024748.

DEDUCTION OF THE FORMULA FOR NPV

The NPV (net present value of an investment) is calculated follows. Imagine a project with an initial investment of a_0 that generates revenue on an annual basis (annuities). Let the annual cash flow be a_i (positive or negative) for year i during a period of n years, which represents the duration of the project, where r is the rate of inflation. It is possible to deduce the present value of the annuity a_2 for the second year with the same purchasing power: $a_2 = a_1/1 + r$. Logically, the value of each monetary unit for the second period will be less than that for the first period, as it will be deflated by the rate of inflation.

Generalising this criteria for each of the n years of the investment's duration and adding the present values of all the annual flows gives the present value of the cash flows for a period of n years:

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + a_4 + a_5 + ... + a_n.$$

And since each year:

$$a_2 = \frac{a_1}{1+r} = a_1 \cdot (1+r)^{-1},\ a_3 = \frac{a_2}{1+r} = \frac{\frac{a_1}{1+r}}{1+r} = \frac{a_1}{(1+r)^2} = a_1 \cdot (1+r)^{-2} ...$$

And so on, until reaching year n: $a_n = a_1 \cdot (1+r)^{-(n-1)}$.

Similarly, for an investment I made in year zero for a duration (t) of n years from year zero, and for which each year from $t = 1$ generates positive or negative monetary flows (+/−) a_1, a_2, a_3… a_n, the net present value of the flow arising from the investment, taking into account the application of a forecast rate of inflation of r, will be:

$$\text{NPV} = -a_0 + a_1(1+r)^{-1} + a_2(1+r)^{-2} + ... + a_n(1+r)^{-n},$$

where a_0 has a negative value, since it is the value of the investment made in the first year.

As a result, the present value of a cash flow of £100 per year for five years, applying the forecasted rate of inflation, will be:

$$100 + 95.23809524 + 90.70294785 + 86.38375985 + 82.27024748$$
$$= 454.59505042.$$

Hence, the present value of this cash flow over five years will be 9.0809899% less than the nominal value of £500.

Imagine a case where a business makes an industrial investment of $I =$ £1,000,000 in facilities with a useful life of 10 years. Here, the lifespan of the project is 10 years, starting from year 0, in which the investment is made. A series of positive profits is forecast for the next 10 years, which are given in the following table:

Year	Flows	Cash Flows	Updated Values
0	a_o	-1,000,000.00	-1,000,000.00
1	a_1	15,000.00	14,285.71
2	a_2	16,000.00	14,512.47
3	a_3	120,000.00	103,660.51
4	a_4	150,000.00	123,405.37
5	a_5	80,000.00	62,682.09
6	a_6	150,000.00	111,932.31
7	a_7	170,000.00	120,815.83
8	a_8	180,000.00	121,831.09
9	a_9	125,000.00	80,576.11
10	a_{10}	250,000.00	153,478.31
		NPV	–92,820.20

The company wishes to know whether, given the present value of the positive profits, it will be able to make a surplus in addition to recovering the investment it has made, bearing in mind that the rate of inflation is forecast to be 5%. Applying the formula for NPV:

$$a_0 + I$$

$$\text{NPV} = -a_0 + a_1(1+r)^{-1} + a_2(1+r)^{-2} + \ldots + a_n(1+r)^{-n} =$$
$$= -1{,}000{,}000 + 15{,}000(1.50)^{-1} + 16{,}000(1.50)^{-2} + 120{,}000(1.50)^{-3} +$$
$$+150{,}000(1.50)^{-4} + 80{,}000(1.50)^{-5} + 150{,}000(1.50)^{-6} + 170{,}000(1.50)^{-7} +$$
$$+180{,}000(1.50)^{-8} + 125{,}000(1.50)^{-9} + 250{,}000(1.50)^{-10} = -£92{,}820.20.$$

Carrying out the calculations gives us the present values indicated in the columns to the right of the table, which give a total negative present value of –92,820.20, meaning that the forecasted profits do not make it possible to recover the capital that is invested, with a loss of£92,820.20!

Another criterion used to measure the profitability of investments is the internal rate of return. This entails calculating an implied interest rate that, when applied to the profits forecast to be made by an investment, makes it possible to recover the full value of the capital that is invested, or rather, its net present value is equal to zero. It involves finding the minimum rate of return for which the investment is advisable. Normally, the internal rate of return for a project is compared with the interest rate or the value of money, and the investment is not recommended if this investment is low. The formula for the internal rate of return (IRR) is given by NPV = 0, hence:

$$\text{NPV} = -a_0 + a_1(1+r)^{-1} + a_2(1+r)^{-2} + ... + a_n(1+r)^{-n} = 0,$$

where $a_0 = I$, initial investment. The solution to the problem can be found in solving for r, which in this case is the internal rate of return. In our example, the result of r that makes it possible to comply with the condition of NPV = 0 for the profits and investment in the previous table is $r = 3.45\%$. (This result that can be calculated using a financial calculator, a spreadsheet, or carrying out the successive iterations by hand.)

How are wages determined?

Wages are determined as a result of the forces of supply (employees) and demand (employers) in the labour market. However, throughout the course of history, various theories have been devised on the mechanisms that regulate wage levels. Classical economists believed that a part of the accumulated capital stock was destined for the payment of workers, and this was referred to as the wage fund. The wage was the result of dividing the fund by the number of hours worked. The problem lay in determining the quantity of this capital stock, or rather the part to be allocated to machinery, tools, buildings and merchandise, and the part allocated to the wage fund. According to the theories of Malthus, the purchasing power of wages tends to subsistence level, since the population grows at a rate far in excess of the production of food, until reaching the limit of the minimum level of subsistence. This theory later came to be referred to as the 'Iron Law of Wages'. Adam Smith also believed in this theory of subsistence wages, but admitted that wages could be increased if there was a corresponding increase in the capital stock that was allocated to them, creating different wage levels between different countries, rich and poor.

It is clear that wage levels depend on the capital stock that is available, and it is also clear that they will be higher when the level of this stock is higher, or rather, as the quantity of capital required per hour worked increases, also depending on the level of technology.

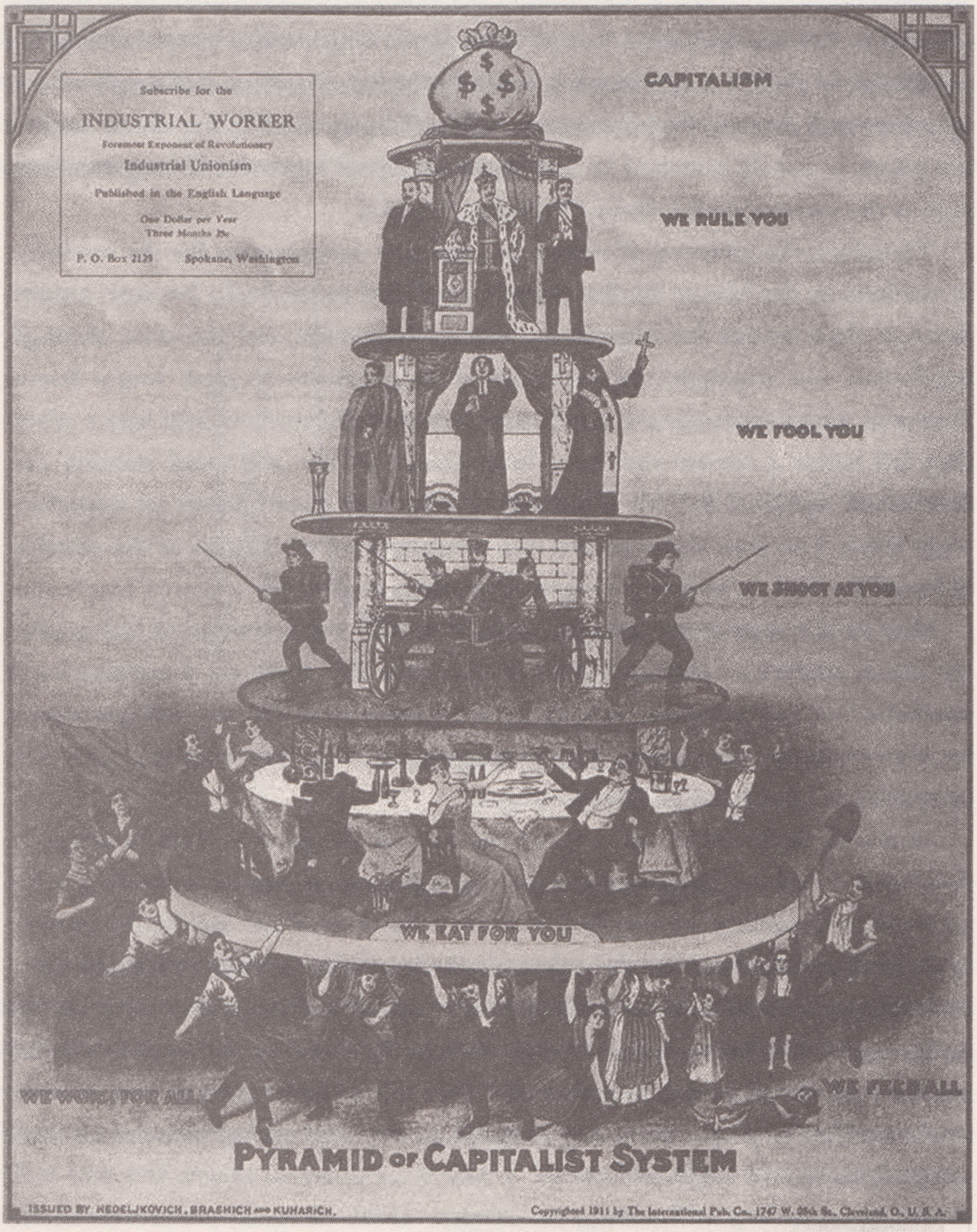
JOHN MAYNARD KEYNES (1883–1946)

One of the problems of greatest concern to economists, beginning with John Maynard Keynes himself, is unemployment. This is due to its damaging effects on society. Unemployment is directly related to the features of the labour market, which sets specific wages for each level of supply and demand. If a market was completely free, with perfect competition, a certain quantity of labour demanded by businesses would establish a level of supply at a wage level determined by a point of equilibrium, hence setting the level below which the supply, or rather the workers, would no longer be willing to work.

However, while Karl Marx believed that capitalism treated labour as if it were just another commodity, the formation of wages has social implications that prevent it from being treated as such – or at least in the eyes of Keynes. He was the first economist to classify employment as a macroeconomic variable and focus on economic analysis from the perspective of demand, in some senses similar to Malthus.

John Maynard Keynes (right) talking to Harry Dexter White of the US Department of the Treasury during the inaugural meeting of the International Monetary Fund in 1946.

In the labour market, both parties find themselves in a highly unequal system. On the one hand, there are the companies, who enjoy a position of power for two reasons: they have an immense quantity of growing economic resources (capital) and, moreover, these are often concentrated in the hands of a few large firms. On the other hand, there are the individual workers with their requirements for


Subscribe for the
INDUSTRIAL WORKER
Foremost Exponent of Revolutionary
Industrial Unionism
Published in the English Language
One Dollar per Year
Three Months 25c
P. O. Box 2129
Spokane, Washington
CAPITALISM
WE RULE YOU
WE FOOL YOU
WE SHOOT AT YOU
WE EAT FOR YOU
WE WORK FOR ALL
WE FEED ALL
PYRAMID OF CAPITALIST SYSTEM
ISSUED BY NEDELJKOVICH, BRASHICH AND KUHARICH.
Copyrighted 1911 by The International Pub. Co., 1747 W. 25th St., Cleveland, O., U. S. A.

A poster from 1911 denouncing the exploitation experienced by workers at the time.

subsistence, the weakness of whose position is evident in the face of the demand of the large companies.

In his book *American Capitalism* (1956), John Kenneth Galbraith analyses how companies tend to concentrate in the formation of large monopolies, referred to as 'technostructures', that play a decisive role in the determination of prices and wages. To counter this enormous power, an opposing force arises spontaneously in society, referred to as a 'compensatory power'. The latter, which tends to re-establish the conditions of perfect competition, is made up of the associations of suppliers, co-operatives and consumer associations, as well as the state itself. In the labour market, this power is mainly exercised by unions, the state and supranational organisations, such as the International Labour Organisation (ILO).

Therefore, the presence of unions and the state compensates for the weakness of workers in the labour market when considered individually, because unions, through their power for negotiation and mobilisation, and the state, through regulation, can attenuate the tendency of corporations to push wages down. However, wages represent the purchasing power of the workers, and hence when they fall below a certain threshold both quality of life and social stability can deteriorate significantly. This is the reason why most governments set minimum wage levels. The state also regulates the conditions of work, such as the number of hours worked, as well as the rights of workers in general.

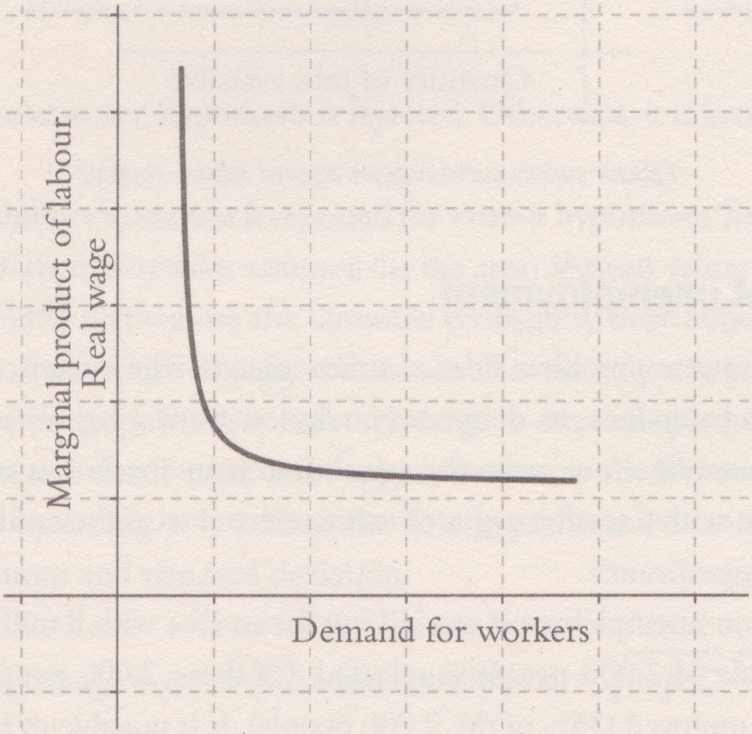
Due to the regulatory activity of the state and the activities of the unions, the labour market is somewhat rigid, in the sense that there are minimum levels established for each professional category as the result of labour laws and agreements between companies and unions. Every company has its organisational structure, with categories of labour and established wages.

In recent years, there has been an increase in the general level of wages in developed countries. The average wage – an indicator of general well-being – has increased considerably. Principally this has been the result of two factors: the increase in the demand for labour and the increase in the level of technology and productivity. Productivity is increased not only by more capital and innovation, but also by increasing the level of education among workers. Furthermore, in conditions of full employment, the reduction in the supply of the active labour force and the increase in the number of people unwilling to work for the wages offered by the demand, can put pressure on wages to rise.

Wages increase until reaching the revenue of the marginal product of labour multiplied by the price of the product, in other words:

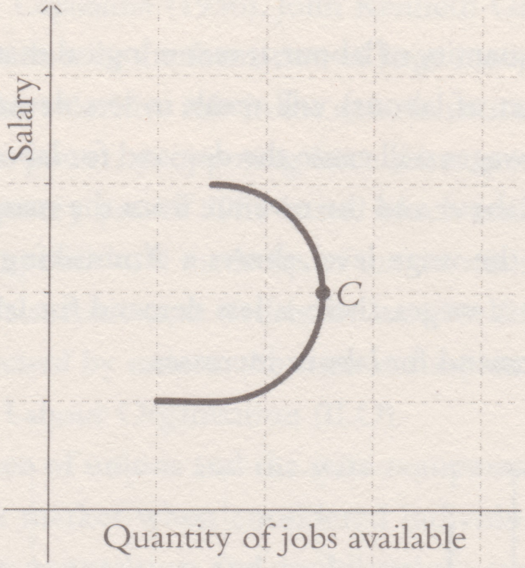
$$\text{Marginal product of labour} \cdot \text{Price of product} =$$
$$\text{Revenue of marginal product of labour} = \text{Wage.}$$

If we consider the quantity of labour, it seems logical that higher wages (revenue of the marginal product of labour) will result in less demand for labour by businesses, whereas falling wages will cause the demand for labour to increase. Drawing the demand curve for labour and the revenue from the marginal product of labour, which corresponds to the wage level, shows a diminishing convexity (see below). In other words at higher wages, there is less demand for labour and, conversely, as salaries decrease the demand for labour increases.



Curve showing wages and the demand for labour.

On the other hand, if we analyse the supply side of the labour market, as the wages offered by the businesses that make up the demand for labour increase, supply also increases. This continues until reaching a determined level of wages where the supply of labour may decrease, because the increase in the purchasing power of workers means their working day will tend to decrease and, as the result of a substitution effect, they will choose to spend more time at leisure. The effect of this substitution is illustrated at point *C* on the labour supply curve below. The level of wages peaks and the supply of labour begins to decrease, tracing a concave path towards the vertical axis, such that there is a maximum labour supply for a given peak level of wages that changes the direction of the curve.



Labour supply curve (wages against labour supply).

The study of unemployment

An extremely common problem in economics concerns the inference of hypotheses. To infer means to deduce, to derive a conclusion from a hypothesis. To accept a statistical inference does not mean the conclusion is an absolute truth, just that it is accepted as true with a small margin of error, referred to as the confidence interval or the level of significance.

In a survey on unemployment carried out for an area with 8 million inhabitants, a random sample of 2,000 people is selected. Of these 2,000 people, 700 declare themselves unemployed (35% of the 2,000 people). Is it possible to infer something about the rates of employment for the area? In other words, is it possible to say that the percentage of unemployment for the area is 35%, 34% or 36%? What we want is a result that is 95% reliable or has a level of significance is 5% ($\alpha = 5/100 = 0.05$).

In order to answer this question, we must formulate a number of hypotheses for the percentage of unemployment in the area. These hypotheses must be taken around 0.35 (35%): 34, 36, 33, 37, 32, 38…, until we find one above and below the 35% that must be rejected. Given that a level of significance of 5% is required to determine if one of the hypotheses should be rejected or not, we must determine if:

The probability that

the number of unemployed (k) in the population (hypothesis) above or below the average $|k-\mu|$ **is greater than or equal to** the number of unemployed (700) in the sample above or below the average $|700-\mu|$

is less than

the level of significance

$(\alpha = 5/100 = 0.05)$.

If this is the case, the hypothesis is rejected. Otherwise, it is accepted.

This probability calculation is repeated for various hypotheses. Each hypotheses is a value p (% of unemployment assumed for the area).Various values of p are selected around 0.35 (35%) and we use the binomial probability distribution, which in this case is suitable since we are dealing with a variable with only two possible responses: yes or no. Since the population is extremely large in our example (n = 8 million) we can use the normal distribution as an approximation to the binomial distribution.

We'll start by taking $p = 0.33$ as the first hypothesis. For this hypothesis, we calculate the mean and standard deviation:

$$\text{average} = \mu = n \cdot p \text{ standard deviation} = \sigma = \sqrt{n \cdot p \cdot (1-p)},$$

which gives

$$\mu = 2000 \cdot 0.33 = 660;\ \sigma = \sqrt{2000 \cdot 0.33 \cdot 0.67} = 21.03.$$

The number of unemployed k in the total population above or below the mean is $|k-\mu|$ (the absolute value of $k-\mu$) and the number of unemployed in the sample above or below the mean is $|700-\mu|$ (the absolute value of $700-\mu$).

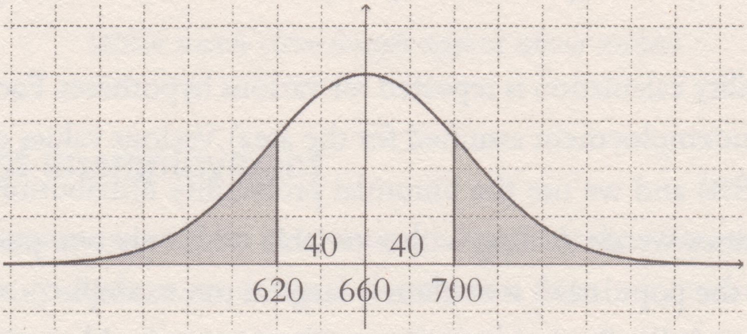
For the hypothesis $p = 0.33$ to be true, the probability according to the binomial law $PB(|k-\mu| \geq |700-\mu|)$ must be less than $\alpha = 0.05$, which is expressed as follows:

$$PB(|k-660| \geq |700-660|) < 0.05.$$

Carrying out the appropriate operations on the inequation gives:

$$|k-660| \geq |700-660| \Rightarrow |k-660| \geq 40 \Rightarrow -40 \geq k-660 \geq +40 \Rightarrow$$
$$\Rightarrow -40+660 \geq k-660+660 \geq +40+660 \Rightarrow 620 \geq k \geq 700.$$

The binomial law approximates the normal law, which is symmetrical with its axis of symmetry on the mean of the distribution, which has a value of $\mu = 660$, for $p = 0.33$. Hence:



$PB(|k-660| \geq 40) = PB(620 \geq k \geq 700) = 2PB(k \geq 700)$, because the two shaded areas are equal.

To approximate a binomial law (*PB* variable *k*, discrete), the following correction must be applied to a normal law (*PN* variable *x*, continuous):

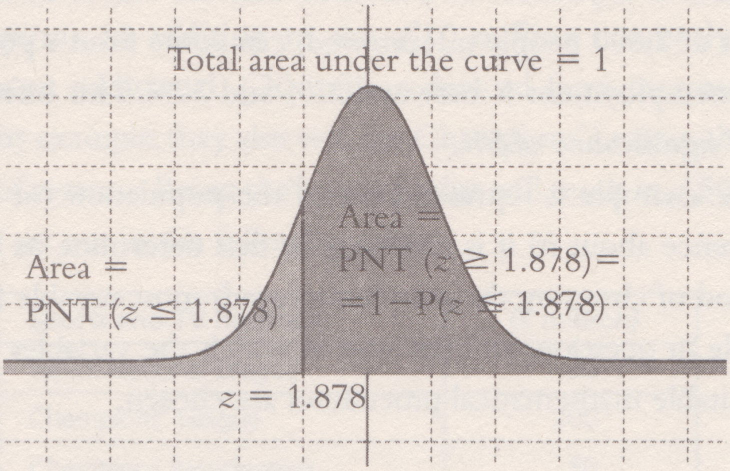
$$\text{PBinominal } (x \geq a) \cong \text{PNormal } (x \geq a - 0.5).$$

Hence, as can be seen from the graph,

$$PB(|k-660| \geq 40) = 2 \cdot PB(k \geq 700) \cong 2 \cdot PN(x \geq 700-0{,}5) = 2 \cdot PN(x \geq 699{,}5).$$

We now apply the change to the normal variable (x) to the standard normal variable (z) to allow us to use the standard tables for the normal distribution. This change of variable is carried out using the formula

$$z = \frac{x-\mu}{\sigma} = \frac{x-660}{21.03}.$$



As can be seen from the graph,

$$PB(|k-660|\geq 40)\cong 2\cdot PN(x\geq 699.5)=2\cdot PNT\left(z\geq\frac{699.5-660}{21.03}\right)=$$

$$=2\cdot PNT(z\geq 1.878)=2\cdot\left[1-PNT(z\leq 1.878)\right].$$

In the tables for the standard normal law, the value $z\leq 1{,}878$ corresponds to a probability of $PNT(z\leq 1.878)=0.96999$ and $PB(|k-660|\geq 40)\cong 2(1-0.96999)=0.0602$, which is greater than 0.05. Since the probability corresponding to the hypothesis of p = 0.33, which is 0.0602, is greater than 0.05, we should NOT reject the hypothesis that the sample represents a population of 8 million people in which 33% are unemployed. Or put another way, we can be sure this area has an unemployment rate of 33% with an error less than 5%. These calculations must be repeated for other hypotheses for p around 35% (p = 0.35), for example, 34%, 36%, 37%, 38%...

With the results we have obtained, it is possible to create a table that solves the problem for each value of the hypothesis p as follows:

(p) Hypothesis	μ	σ	z	From the SNP (z) table	Calculation of $PB\vert k-\mu\vert\geq\vert\vert 700-\mu\vert$	Accept Hypothesis? $PB < \alpha$?
0.32 (32%)	640	20.8614477	2.85	0.9978	0.0044	No
0.33 (33%)	660	21.028552	1.88	0.9699	0.0602	Yes
0.34 (34%)	680	21.1849003	0.92	0.8212	0.3576	Yes
0.35 (35%)	700	21.330729	−0.02	0.5	1	Yes
0.36 (36%)	720	21.4662526	0.91	0.8186	0.3628	Yes
0.37 (37%)	740	21.5916651	1.83	0.9664	0.0672	Yes
0.38 (38%)	760	21.7071417	2.74	0.9969	0.0062	No

From this table it is possible to conclude that the values of the sample (700 unemployed out of 2,000 people interviewed) can come from a population whose percentage of unemployment is between 33% and 37%. This information has an error or level of significance of 5%.

To guarantee a sample is representative of the population (or rather, to make a statistical inference about it) it is necessary to first determine its size, the type of sampling (method of choosing the people surveyed) most suitable for studying the statistical variable in question, and the way in which the variables to be estimated and the most suitable mathematical procedures are chosen.

Is it profitable to sign a star player?

It is easy to feel aggrieved when we hear that a footballer or any other sports star has been signed for an astronomical sum. What are the economic reasons that can justify such a considerable investment? Remember it is necessary to add the player's annual salary to the transfer fee, and this can be in the millions.

The formulation of the transaction is highly complex. If a sports team signs a star, it is possible they will play better and win more games. If the team wins, the club will gain more supporters who will come to watch, the number of season ticket holders or club members will increase, with the consequent rise in membership fees, more money will be taken for the sale of tickets at the grounds, advertising revenue will increase, more football strips and sporting merchandise with the club's colours will be sold, the value of the rights to transmit their games on television will increase and, if they win competitions, they will receive substantial sums of money from the national and international federations that organise them. The problem is then, to determine if the high costs of signing the player and paying his or her salary are compensated by the appearance of these positive results.

To illustrate this, we'll use the following example: A football team with 22 players has not won a competition for many years. To make things simpler, imagine that the total expenditure of the club is limited to the cost of the players' wages, which average £3 million per year, or rather the total annual cost of the playing staff is £66 million. On the other hand, the club's revenue comes from two sources: it receives £10 million a year from ticket sales for playing in the league and payments for the rights to show their games on television, and it has an advertising contract that pays £5 million a year. In total, the club has annual revenues of £15 million, compared to annual expenditure of £66 million, which makes a loss of £51 million a year.

There is considerable potential to attract more revenue by winning competitions. As league champions, in addition to the £10 million currently received, the team would earn an additional £30 million derived from increased ticket sales. Furthermore if, for example, they also win the Champions' League, this would result in £200 million of revenue. Being the losing finalist will result in £50 million. These positive flows are set out in the following table:

Sports and TV Revenue	(£ million)
Top of National League	30
Champions' League	200
Champions Participation	50
League Participation	10

The revenue from the Champions' League and winning the national league is on top of the £10 million the team normally earns without winning any competition. In addition to this, there is also advertising revenue, which will double from its current value of £5 million a year if the team wins the league. Winning the Champions' League will also result in additional advertising contracts with a value of £40 million. If the team just participates in the competition, without reaching the final, advertising revenue will increase by £10 million. Finally, the best-case scenario, in which they win both the Champions' League and the home league in the same year, will result in increased advertising revenue of £70 million. To summarise, the potential levels of advertising revenue depending on their performance are as follows:

Advertising Revenue	(£ million)
Play in League	5
Win League	10
Play in Champions' League	10
Win Champions League	40
Win League and Champions	70

During the last five years, the club has lost £255 million and its directors have decided to sign a star player to improve their sporting fortunes and turn around the club's finances. The transfer fee is £100 million and the player is to be paid a salary of £7 million a year, for a five-year contract. The transfer goes smoothly, although not without attracting scandal in the sports media. The issue is whether it is possible to determine if the duration of the contract and the cost of the transaction for signing the star makes it profitable based on the team's results over the coming years.

In the years after signing the star, the club's fortunes have improved considerably. They won nothing the first year, but finished towards the top of the classification, and the following year the club played in the Champions' League. During the second year, in addition to playing in Europe, they won the home league. In the third year, they did not win the league but won the Champions' League instead. In the fourth year they won the national league and played in the Champions' League, and in the final year of the contract, they only played in both competitions, without winning either. This gives a total of three wins in five years: two home leagues and one Champions' League.

At the end of the player's contract, the directors of the club want to know the profit from the five-year transaction, and they draw up a table to compare the results with the incorporation of the new star with those they would have obtained if they had continued under the previous circumstances without spending a total of £135 million on the signing (a £100 million transfer fee plus £35 million paid in wages). The results of the comparison are as follows:

	With Star	Without Star	Competitions Won
Costs			
Price of Signing	100	0	
Player Price: 7 × 5	35	0	
Staff Costs	330	330	
Total Costs	465	330	
Revenue			
Tickets for Games and TV			
Year 1	10	10	0 League, 0 Champions
Year 2	90	10	1 League, Play in Champions
Year 3	210	10	0 League, 1 Champions
Year 4	90	10	1 League, Play in Champions
Year 5	60	10	0 League, Play in Champions
Total Tickets and TV	**460**	**50**	
Advertising			
Year 1	5	5	0 League, 0 Champions
Year 2	20	5	1 League, Play in Champions
Year 3	45	5	0 League, 1 Champions
Year 4	20	5	1 League, Play in Champions
Year 5	15	5	0 League, Play in Champions
Total Advertising	105	25	
Total Revenue	565	75	
Profit	100	–255	

Cash flows with or without signing the star (5 years and £ million).

As can be seen, at the end of the contract, in addition to recovering the £135 million transaction cost, the club has made a profit of £100 million. However, the directors decide to calculate the net present value of this positive cash flow, for which the value of money is 4%. They also want to know the internal rate of return of the cash flows in order to evaluate the return of the operation in line with various criteria. The results obtained are as follows:

	Total Revenue	Total Costs	Profit
Year 1	15	173	–158
Year 2	110	73	37
Year 3	255	73	182
Year 4	110	73	37
Year 5	75	73	2
Total	565	465	100
NPV	80		
Interest Rate of 4%	0.04		
IRR	28%		

Cash flows from signing a star.

It can be seen that the NPV (net present value) is positive (£80 million), and furthermore the internal rate of return for the sporting project in which the star was signed is highly positive (28%). However, there is still a fundamental flaw in this analysis, since it fails to take into account the possibility of not winning anything, a risk that must be considered (most teams do not win). The following table has been drawn up to calculate the profits that would have been made signing the star player but without winning a competition.

	Total Revenue	Total Costs	Profit
Year 1	15	173	–158
Year 2	15	73	–58
Year 3	15	73	–58
Year 4	15	73	–58
Year 5	15	73	–58
Total	75	465	–390

Cash flows for signing the star but without winning a competition.

As we can see, in this scenario the profits would have worsened considerably with respect to previous performance. The loss of £390 million would have represented a 50.94% increase on the losses for the previous five-year period. Hence, when it comes to calculating the profitability of any business transaction, it is important to evaluate not only the real results and forecast them to make future predictions, but also to evaluate future scenarios and take into account the risks and threats that may arise, applying the calculation of probabilities with which certain events may occur, regardless of whether they are opportunities or threats.

Chapter 5

How to Buy at the Lowest Possible Price: the Market

The market began its life as the place where buyers and sellers met. Manufacturers and merchants offered their products and citizens bought the goods they needed. It goes without saying that the sellers sought to maximise their profit and ensure the greatest possible margin between the price at which their products were sold and the price at which they were produced. For their part, purchasers sought to buy their goods at the lowest possible price. Sellers knew perfectly well that if they put their prices up, the number of sales would drop, and if over the course of the day they had failed to sell their goods and thus reduced their prices, the goods would sell quicker. This sounding-out of buyers and sellers eventually came to establish relatively stable prices that oscillated depending on weather patterns and periods of scarcity.

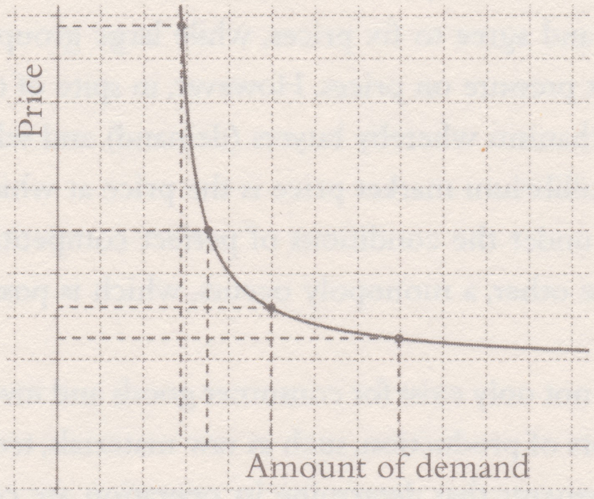
Supply and demand

The market is as old as humanity, but it is now a highly evolved and complex mechanism. It is possible to produce products thousands of miles away, and for these to pass through many intermediary transactions and sales. A small group of suppliers may come together and agree to fix prices, while large groups of consumers may also team up to exert pressure on prices. However, in spite of this, the market continues to be the mechanism whereby buyers (demand) and sellers (suppliers) meet to set prices. The equilibrium market price is the price at which buyers and sellers reach an agreement under the conditions of perfect competition. If one party has more power than the other, a monopoly occurs, which is possible on both supply and demand sides.

The market does not only exist for consumer goods and investment and services, but also for the factors of production, such as raw materials, technology and labour. The fundamental elements that determine its operation are consumer preferences (their tastes and how much they are willing to pay) and the production costs of the supply – plus commercial strategies for maximising profits.

The market is the traditional meeting place for buyers and sellers although today the term has a much broader meaning, to the point of being one of the main characteristics of the capitalist economy. This photo shows the central market in the US city of Lancaster, Pennsylvania.

Demand is studied using historic statistical tables for the level of demand of a good, in line with the changes in prices. A specific level of demand corresponds to each price. The demand curve is graphically represented by a descending hyperbola, with prices represented on the horizontal axis and the levels of demand on the vertical axis.



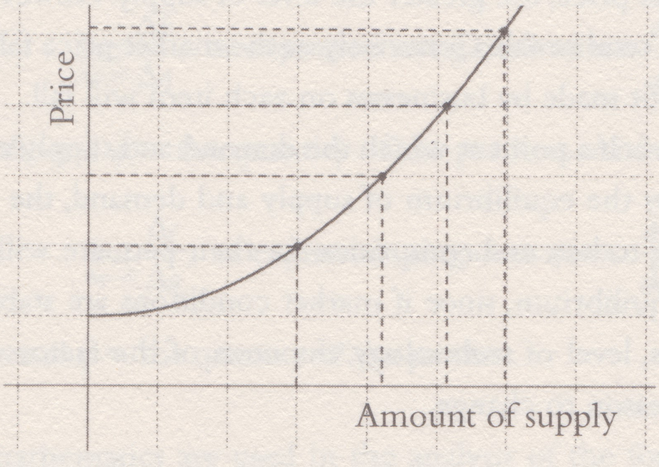
Demand curve.

The demand curve is not static, but instead moves in line with economic conditions. Hence if the wage levels of consumers rise, it is clear there will be more demand for goods at the same price, causing the demand curve to shift to the right. The same thing happens when the number of consumers increases. On the other hand, if wage levels fall as the result of an economic crisis, the demand curve will shift to the left.

Hence, we must distinguish between movements along the demand curve, which are only related to the levels of demand for the goods and the corresponding prices, with the other variables that affect the market remaining constant (i.e. wage levels and number of individuals), and those related to increases in demand caused by changes in the other variables that affect the formation of prices and cause the demand curve to shift to the left or right.

The supply curve represents the quantities of a product companies are willing to offer at certain prices, that allow them to make a profit.

Supply is also represented on a graph in which prices appear on the vertical axis and the corresponding levels of supply on the horizontal axis. The graphical representation takes the shape of an ascending parabola, which starts at a minimum price below which there is no supply, and which is below the cost of production.



Supply curve.

An extremely important aspect now arises in the study of supply: the costs of production. The lower production costs are with respect to price, the higher the profit that will be made by the company and hence it will be willing to supply the market with more of the product.

Production costs are significantly affected by the costs of the factors of production, such as labour, energy and machinery, the combination of which is determined by the level of technology used by the company. Hence, as it consumes less energy and labour, machinery will be used more efficiently, a phenomenon that is rooted in technological progress. This also entails the search for substitute factors that increase the efficiency and productivity of a company.

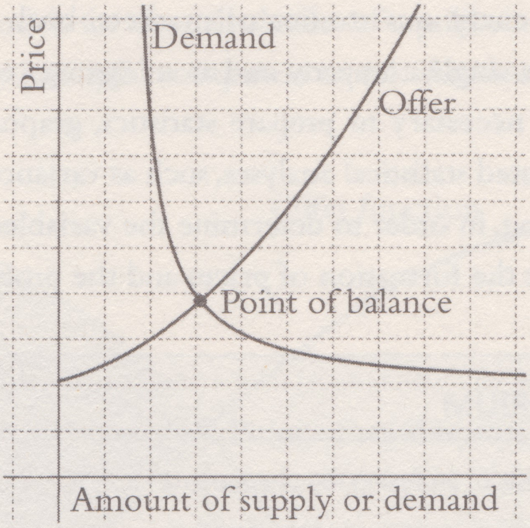
This is how the production process has been automated, replacing human labour with automated production lines, accompanied by the subsequent wage savings; materials are replaced by others that are stronger, more reliable and cheaper.

Just like the demand curve, the supply curve can also shift. If market prices change, the level of supply will continue to follow the supply curve. However, if the technology changes or the number of competitors increases, the total supply may increase, causing the supply curve to also shift to the right.

The theory of market equilibrium

The higher the market price, the greater the level of supply. However when prices are high, demand will tend to fall. Conversely, as the market price falls, demand will rise, although the profit made by businesses on each item will fall.

The process will reach a point at which the demand and supply curves intersect, and this point denotes the equilibrium of supply and demand, the price at which consumers are willing to buy and companies, for their part, are willing to sell. It is called the point of equilibrium, since if market conditions are stable (wage levels, consumer preferences, level of technology, the costs of the factors of production, etc.) it will have no reason to change.



Supply and demand curves.

However, the point of equilibrium can also be influenced by movements of the supply and demand curves – by variations in the level of technology and increasing competition on the supply side, as well as by variations in wage levels and the number of consumers, or even just their preferences and tastes, on the demand side. On the supply side, the point of equilibrium for the market will occur when the marginal cost (the increase in production costs generated by producing an extra unit of the product) is equal to the marginal profit (the profit produced by the new unit that is sold).

Therefore, the marginal profit is:

$$\lim_{\Delta q \to 0} \frac{\Delta b}{\Delta q}; \text{ or expressed using derivatives: } \frac{db}{dq},$$

where b = profit and q = the quantity of the product. And the marginal cost is:

$$\lim_{\Delta q \to 0} \frac{\Delta c}{\Delta q}; \text{ or expressed using derivatives: } \frac{dc}{dq},$$

where b = cost and q = the quantity of the product.

Powerful mathematics are used in the analysis of the formation of prices and markets. First of all, it is necessary to prepare statistics for sales, levels of production and prices, in addition to analysing the prices at which supply and demand are in equilibrium. In other words, determine how much demand a product will generate, given its price, and given the cost curves determine how much profit will be made by a company (or the sector of the economy) when the study is carried out at a macroeconomic level.

It is clear that to carry out studies of markets, both at an individual level (microeconomics) for a single company and at an aggregate level for all companies in a given sector, it is necessary to prepare statistics, graphical analyses and other types of more complicated statistical analyses, such as variance analysis, multivariable analysis and curve fitting, in order to determine the variables other than supply and demand that will affect the formation of prices and the profits made by companies.

GENERAL EQUILIBRIUM

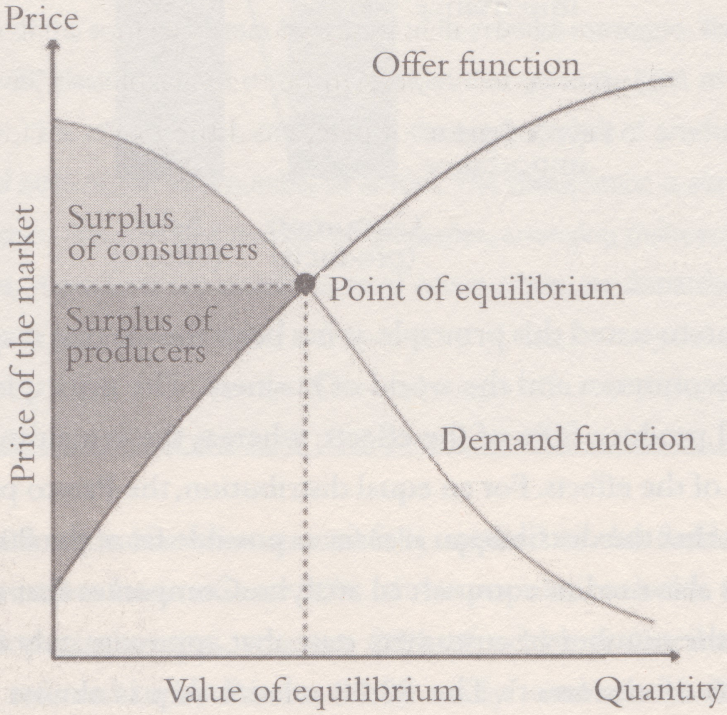
Léon Walras, the French economist responsible for providing the mathematical foundations of general equilibrium.

General equilibrium consists of analysing the phenomena of an economy when all of its sectors are considered simultaneously, such that all direct or indirect impacts and the relationships between markets are included. Léon Walras (1834–1910) was the first to formulate general equilibrium using mathematical notation..

He believed the company was the centre of the economy and studied its activity in terms of competition between agents. He also observed an interdependence between all markets in an economy: product markets (goods and services) and factors of production (land, labour and capital). He analysed the procedures for simultaneously setting prices and quantities and formulated the economic problem of general equilibrium, or rather the stability of equilibria in the market. The Italian Vilfredo Pareto and the Austrian Carl Menger continued his work and are regarded as the fathers of the neoclassic marginalist school of economics.

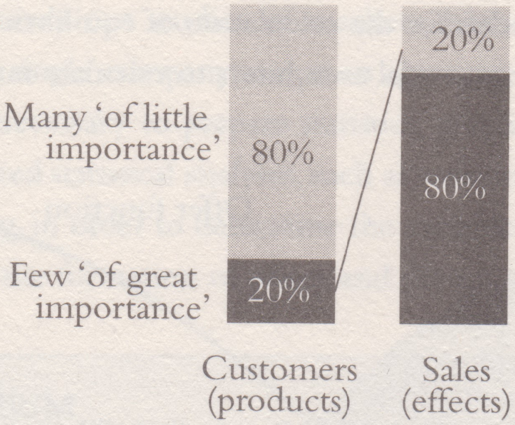
The theory of general equilibrium is studied in microeconomics (focused at the level of the company). It is a global explanation of the behaviour of production, consumption and the setting of prices in an economy. General equilibrium attempts to provide an explanation that goes from the specific to the general (*bottom-up*), beginning with the markets and their individual agents. In contrast, the macroeconomy (the aggregation of microeconomic variables) is based on a vision that goes from the general to the specific (*top-down*), and analysis begins with the largest components.

Modern general models for the calculation of equilibrium are extremely complex and require highly powerful computers to calculate numerical solutions.

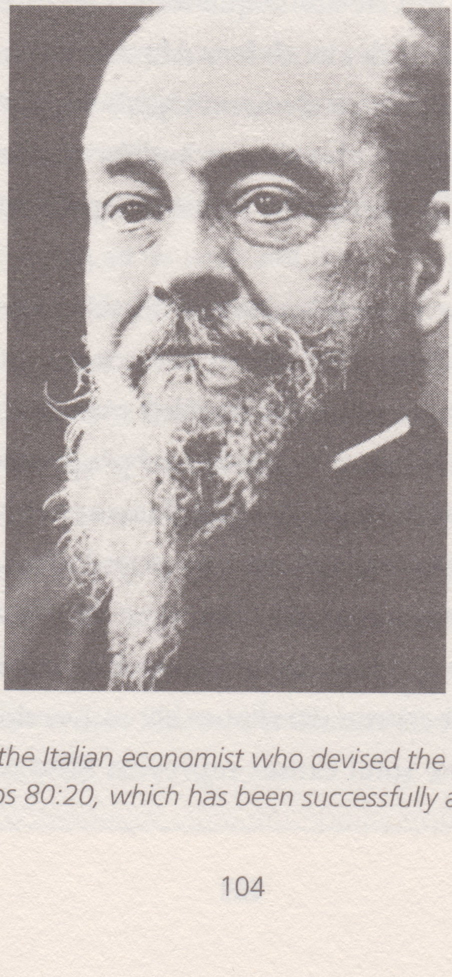


In the Walras model the functions for the prices for the level of supply and demand are represented by interrelated equations. The three variables are price, the level of supply and the level of demand. It is necessary to calculate two unknowns: price, and the equilibrium at which demand equals supply. For the supply curve or function, price increases in line with quantity and for the demand curve or function, price decreases in line with quantity. The point of equilibrium is the point at which the two curves meet.

On the other hand, the principle devised by Vilfredo Pareto (1848–1923) is based on empirical knowledge. The Italian economist and philosopher observed that in his society, people were naturally divided into the 'vital few' and the 'trivial many'. This led him to establish two groups of proportions 80:20, such that the minor group, composed of 20% of the population, held 80% of something, and the major group, composed of 80% of the population, held 20%. The figures are arbitrary; they are not exact and may vary. However they describe a real phenomenon and as such can be adapted to individual cases. Bear in mind that both the distribution of effects and their possible causes does not constitute a linear process, but that 20% of the total causes represent 80% of the effects.



Ever since Pareto stated this principle, it has been shown that it applies to a range of situations in economics and the world of business. The principle states that 20% of an action will produce 80% of the effects, whereas the remaining 80% will only give rise to 20% of the effects. For an equal distribution, the Pareto principle must be minimised, such that the distribution is as far as possible from the 80:20 proportions. This principle is also used in commercial analysis. Companies that analyse turnover with respect to the number of customers state that approximately 80% of turnover depends on 20% of customers. The 80:20 relationship is almost never observed exactly, but the imbalance between sales and the number of customers often occurs.



Vilfredo Pareto, the Italian economist who devised the principle of the two proportional groups 80:20, which has been successfully applied in many fields.

THE ABC CURVE OF A WAREHOUSE

The Pareto principle is also used to analyse the selection or range of products sold by a business, where 80% of the turnover comes from 20% of its product catalogue. This has given rise to the so-called *ABC* classification: *A* = 20% of articles that generate 80% of the warehouse transactions; *B* = 15% of total stock and 30% of the total number of articles; and *C* = 5% of total stock and 50% of the total number of articles. The classification is also used to group items in a warehouse into a limited number of categories, arranging them according to their availability. Hence, *A* products are located close to where orders are dispatched to save time when moving goods around.

The Pareto principle has been successfully applied in the fields of politics and economics. It describes how around 20% of the population holds 80% of political power and economic resources, whereas 20% of wealth is distributed among the remaining 80%, referred to by Pareto as 'the many', and who hold little political influence. In general terms, the same thing happens with the distribution of natural resources and global wealth.

Elasticity of demand

Economic resources have a certain scarcity and there is a cost attached to obtaining them that must obviously be measured in physical and monetary units. Once produced at a given cost, they must be sold. The prices at which products are sold are determined in line with demand and, if they are greater than the costs of production, will result in a profit. This gives rise to three types of requirements affecting decisions made by entrepreneurs:

1. Forecast the evolution of demand and hence changes in retail prices.
2. Attempt to minimise production costs in order to maximise profit, and determine the threshold of profitability below which it is no longer acceptable to continue producing the good or service in question.
3. Prepare a mathematical model that makes it possible to determine the condition of market equilibrium and see how prices are formed in line with the levels of supply and demand.

Hence, it is necessary to study how demand varies in line with changes in the price of the supply. This is done by defining a relationship between the change in demand and the change in prices. This is referred to as the elasticity of demand and expressed by the following formula:

$$\text{Elasticity of demand} = \Delta q \;/\; \Delta p,$$

where Δq is the variation in the level of demand for a good and Δp is the variation in its price.

It can also be expressed in terms of percentage variations of the level of demand over the percentage variation in price:

$$\text{Elasticity of demand} = \frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}}\text{, the same as: } E=\frac{dq}{dp}\cdot\frac{p}{q}.$$

Hence, the derivative of q with respect to p, $\frac{dq}{dp}$ multiplied by $\frac{p}{q}$.

The elasticity of demand is measured in unitary terms; hence it measures how much the level of demand for a product will vary in line with an increase in price of one monetary unit. If there is a large change in demand for a small change in prices, demand is said to be elastic. If prices increase considerably, demand will fall proportionally until consumers are required to purchase other substitute goods at a lower price. When demand is elastic, its elasticity is greater than one. In other words, for a price increase of one unit, demand will increase by multiple units. Conversely, if demand hardly varies with respect to variations in prices, demand is said to be inelastic. Under these circumstances, the elasticity of demand is less than one.

Goods to satisfy basic needs, such as food, clothes and footwear, often have inelastic demand, whereas the demand for luxury goods is often more elastic, since as their price increases, they are substituted for cheaper goods, whereas if the price falls, demand quickly shoots up.

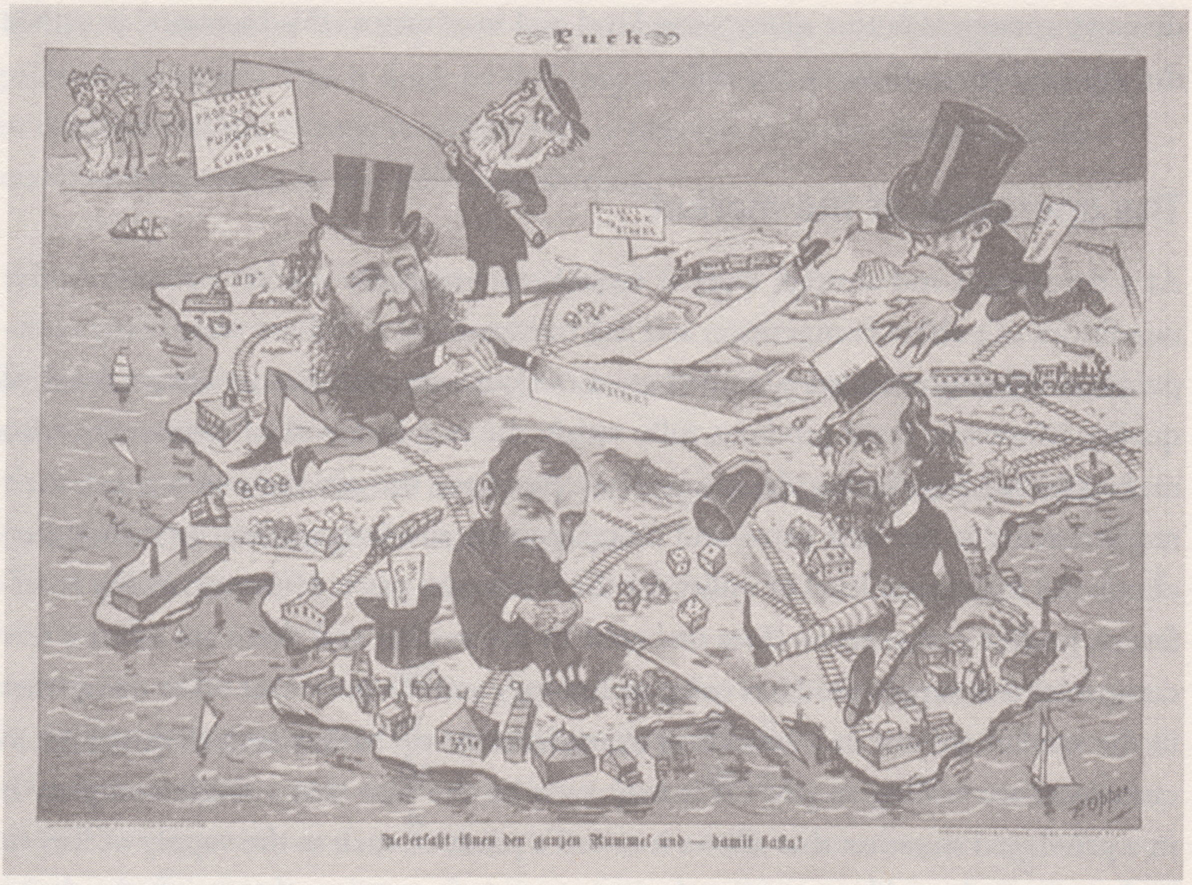
The formation of prices is also influenced by the wage levels of consumers, since as wages rise they are able to purchase more goods. Hence, the elasticity of demand corresponds to a given wage level and may vary as this increases, or rather the demand curve shifts to the right.

The operation of the market

At a theoretical and abstract level, it would seem as if the market operates under conditions of perfect competition and full transparency, as if there were many companies supplying products obtained at minimum cost and the maximum level of quality. Moreover these products fully satisfy the requirements of consumers who make rational purchasing decisions. They purchase the products that meet their requirements at the best price and comparatively the best quality, such that they obtain the maximum level of utility for the minimum price. However, reality is different, perfect competition does not exist in the market and there are often a few companies that supply products using technology and at a price that other companies in the market are not in a condition to replicate. As such, they set prices as high as customers are willing to pay. Or it may even be the case that a single company is in a privileged situation that allows it to sell a product, such as the supply of fuel or high-tech products like precision devices or aircraft, at prices that satisfy the desire for businesses to make a profit rather than the interests of the consumer. In other cases, companies launch innovative products or services onto the market, of which the consumer is unaware, since they did not previously exist. Advertising campaigns are used to stimulate demand for the new product, and the consumer responds compulsively (even irrationally), buying a product they do not in fact need, although they have been convinced they do, paying a high price.

There are also a few companies that enjoy privileged positions, such as administrative concessions (e.g. water supplies, electricity and public transport), allowing them to impose high prices on consumers. Without state intervention, monopolies (when a single company dominates the market) can often be highly damaging to consumers, although in some countries it may be the state itself that holds the monopoly. On other occasions, when the supply side is composed of a small number of companies (an oligopoly), these may reach agreements about the prices that should be offered to consumers, resulting in an effective monopoly – absolute control of the market in favour of companies and at the expense of consumers. Oligopolies (or cartels) are common in the automotive and insurance sectors.

Monopoly of demand (monopsony) can also occur, in addition to monopoly of supply, whereby there is a single purchasing company and a range of suppliers. Under


Puck
Ueberlaßt ihnen den ganzen Rummel und — damit baſta!

In this sketch from around 1885, the US illustrator Frederick Burr Opper characterises the distribution of the country carried out by monopolistic millionaires.

these circumstances, the purchaser can impose harsh conditions on suppliers, such as delivery times, heavy discounts on prices, volume discounts, conditions for displaying products, payment terms, etc. One example of monopsony is large retailers imposing heavy discounts on their suppliers. These take the form of volume discounts, payment terms whereby goods cannot be invoiced until they are sold, payment terms in excess of 30 days and compensation for exposure of products. In this last case the big retailer discounts amounts from their suppliers' invoices because they believe the exposure of their products on the supermarket shelves is advertising, which must be paid for. Clearly, when a company wishes to trade with a large retailer, it must sign a supply contract with extremely harsh conditions imposed by dint of the monopsony.

Another factor that influences the final price of products is the existence of intermediaries or 'middlemen', who charge small mark-ups on the initial price, or commissions based on sales figures. The existence of many such channels in the sale of a product can result in considerable increases in the final price that must be paid by the consumer, with certain intermediaries being able to achieve near or total monopolistic positions. This situation often occurs in distribution channels for agricultural products, which leads to the large difference between the 'farm gate' prices at which goods are sold by producers and the cost charged to the end consumer.

Another way in which markets can be organised is through auctions, which make use of forums in which certain products are sold to the highest bidder. There are various types of auction, the most common of which is the ascending price auction (known as an English auction), in which a specific article is offered at a minimum starting price, which is increased by potential buyers until being sold to the highest bidder. This method of sale, which was once carried out in person, is now a common practice on the Internet.

Another type of auction is the descending auction, which starts from a high price, lowering the value until a buyer stops the descent (the so-called Dutch auction).

INVOICE DISCOUNTING

Commercial transactions can be executed in many different ways, especially when it comes to payment term. One is invoice discounting, whereby a company receives an advance payment for an invoice issued to a customer from a bank or similar. The bank will discount a sum of interest for the advance payment, a system referred to as invoice discounting, and the discount can be applied to IOUs, notes payable or simply the endorsement of invoices prior to their due date. If the value of the asset discounted by a bank for a company has a value of V_n and is due on an expiry date *n* to a customer who has accepted the debt, the bank will advance a sum less than V_0 and will apply an annual interest rate *i* for the transaction (adjusted to the period of time *n* for which the advanced payment is received, giving an interest rate i_e), such that the interest charged by the bank will be:

$$D = V_n \cdot i_e.$$

And the value received by the company for the advanced payment V_0 will be:

$$V_0 = V_n - V_n \cdot i_e = V_n(1 - i_e) = V_n - D;\ V_0 < V_n,$$

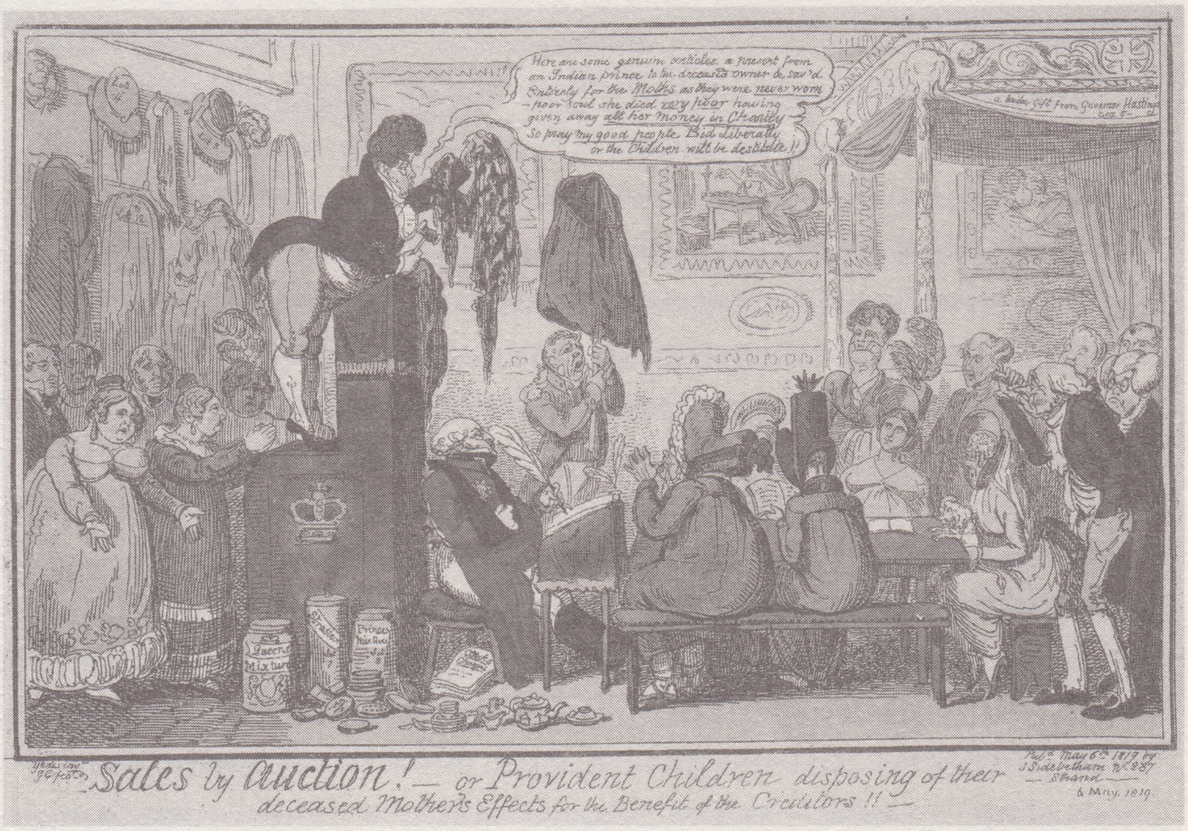
where V_0 is the advanced payment made by the bank; V_n, the value of the discounted bill; *i*, the annual interest rate applied by the bank (i_e, the effective interest rate applied in line with the time that has passed; *n*, the expiry period) and *D*, the value of the interest (= $V_n \cdot i_e$).
It is the same formula as the one used for simple interest but with an advanced charge.

Finally, in an American auction, potential purchasers are required to submit their secret bids in advance.

Auctions can also be used for purchasing, and this is the method normally used by public administrations when they wish to award a project or purchase an asset. These auctions use sealed bids and are awarded to the most economical bidder. A similar method is the tender, in which in addition to price, other factors that may make a bid more valuable are also taken into account.

There are currently many types of auctions, depending on the sector, origin and type of product. Hence there are customs auctions, in which confiscated goods are auctioned; property auctions, in which financial entities auction property that has been seized; judicial auctions, which are of a similar nature; Internet auctions, which are carried out electronically; and collectors' auctions for works of art, curios and antiques.

Auctions are a highly transparent system for setting prices and organising markets, since supply and demand meet in perfect freedom. This is why the system


Here are some genuine articles, a present from an Indian prince to the deceased owner & sav'd Entirely for the Moths as they were never worn - poor soul she died very poor having given away all her money in Charity So pray my good people Bid Liberally or the Children will be destitute !!
a birth gift from Governor Hastings
Queen's Mixture
Sales by auction! — or Provident Children disposing of their deceased Mother's Effects for the Benefit of the Creditors !! —
Pub.d May 6th 1819 by J. Sidebetham No. 287 — Strand —
6 May. 1819.

An English illustration from 1819 representing a scandalous auction in which the impecunious Prince Regent (later to become King George IV) sold various personal effects belonging to his recently deceased mother.

is also used for more complicated markets, such as the foreign exchange market, as it is a highly effective instrument for setting exchange rates, assigning issues of state securities or treasury bills, and allocating the issue of fixed-rate securities. The stock market uses the auction as the system to set the initial prices at the start of trading sessions, or for the prices of securities at the end of the session.

A market study

We will now analyse a market study based on the following scenario. It has been calculated that a brand of detergent is used by 26% of households; a survey is carried out on its use on a sample of 12 households, and we wish to calculate the probability that the number of users of the brand in the sample is between 6 and 9. Hence, we are dealing with a variable that has just two values (if the detergent is used or not), with an estimated probability of 26% and 74% (100–26). Using the laws of combinatorics, we can deduce the probability $PB(k)$ that there are k users of the brand in a sample of n, hence:

$$PB(k)=\binom{n}{k}p^k(1-p)^{n-k},$$

where $p = 0.26$ and $(1 - p) = 1 - 0.26 = 0.74$. The formula corresponds to the binomial probability distribution (or function), discovered by Jakob Bernoulli in the 17th century. On the other hand,

$$\binom{n}{k}=C_n^k=\frac{n!}{k!(n-k)!}$$

denotes the number of unique combinations that can be obtained from n objects in groups of k.

The probability that one or more of the possible events occurs is calculated by adding the individual probabilities for each of these, provided they are independent (they cannot occur at the same time).

DIFFERENT WAYS OF GROUPING OBJECTS

The combinations of *n* objects in groups of *k* objects (with $k < n$) refers to all groups of *k* in which two groups are regarded as different if at least one object is different, regardless of the order in which they are arranged. In the case of the number of combinations of three groups of four elements, the number of possible groups is indicated by C_4^3 and is calculated using the formula below.

$$C_4^3 = \frac{4!}{3!(4-3)!} = \frac{4\cdot 3\cdot 2\cdot 1}{3\cdot 2\cdot 1\cdot 1} = 4.$$

Four different groups can be formed. We can see that if the four elements to be combined are the letters A, B, C, D, the four groups of combinations will be ABC, ABD, ACD, BCD.

Other types of groupings of objects are also widely used in discrete mathematics, including variations and permutations.

The variations of *m* objects in groups of *n* V_m^n means that two groups are regarded as different if they differ by one element or have the same elements in a different order. The number of variations of groups of three for four elements (p, q, r, s) is:

pqr, pqs, prq, psq, prs, psr,
qrp, qpr, qps, qsp, qrs, qsr,
rps, rsp, rpq, rqp, rsq, rqs,
spq, sqp, sqr, srq, spr, srp.

The number of groups is calculated using $V_m^n = m\cdot(m-1)\cdot(m-2)\ \dots\ (m-n+1)$; in this case:

$$V_4^3 = 4\cdot(4-1)\ \dots\ (4-3+1) = 4\cdot 3\cdot 2 = 24 \text{ groups.}$$

Permutations are variations in which each group contains all elements, or rather $m = n$. The permutations of three elements (M, N, P) are the variations of three and the groups are MNP, MPN, NMP, NPM, PMN, PNM. The number of groups is calculated using the following formula:

$$P_n = n\,(n-1)\,(n-2)\dots(n-n+1) = n\,(n-1)\,(n-2)\dots 3.2.1 = n!$$

In this case, $P_3 = 3(3-1)(3-2) = 3! = 6$ groups.

In our example, the probability that six users appear in our sample is:

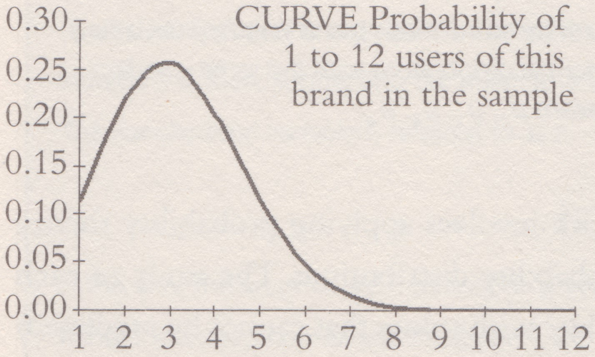
$$PB(6)=\binom{12}{6}0.26^{6}\cdot 0.74^{6}=0.0468708102.$$

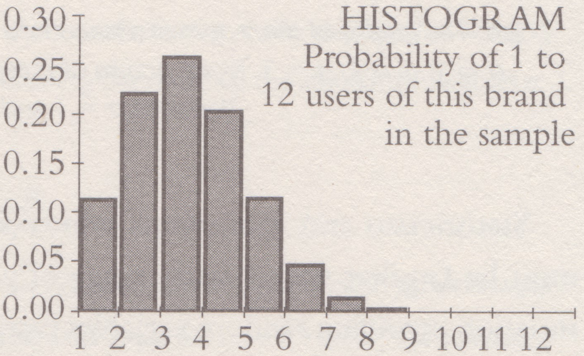
Applying the formula above yields the following table, calculated in Excel, of the values of *PB*(1) to *PB*(12).

k	1	2	3	4	5	6	7	8	9	10	11	12
PB(*k*)	0.1136850903	0.2196887556	0.2572931372	0.2034006557	0.1143441524	0.046870801	0.0141155309	0.0030996943	0.0004840363	0.0000510200	0.0000032593	0.0000000954

The graph for the probability distribution is presented in two ways. On the right (below) is the histogram of the sample size and to the left, a continuous function with the formula

$$PB(k)=\binom{n}{k}p^{k}(1-p)^{n-k}.$$





And the requested probability, corresponding to the values between 6 and 9 is:

$$\begin{aligned}&PB(6\le x\le 9)=PB(6)+PB(7)+PB(8)+PB(9)=\\&=0.0468708102+0.0141155039+0.0030996943+0.0004840363=\\&=0.0645700627=6.46\%\end{aligned}$$

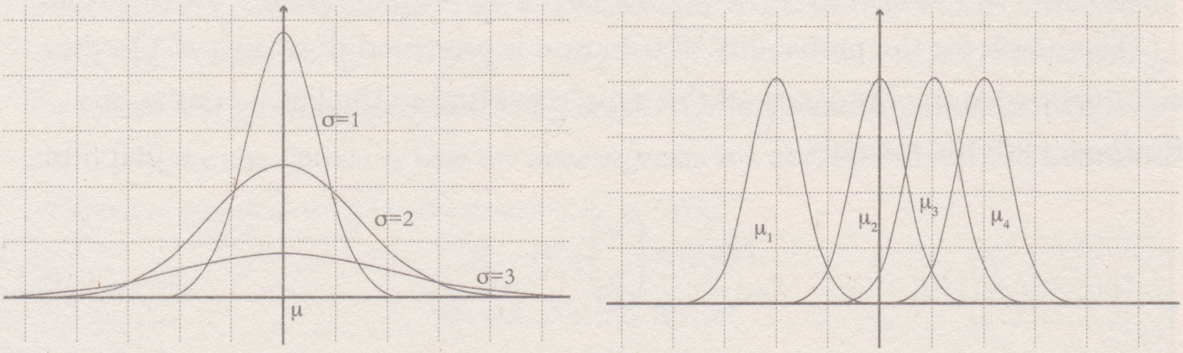
The measures of the central tendency and dispersion of a binomial probability distribution are calculated using the following expressions:

$$\text{mean} = \mu = n \cdot p; \text{standard deviation} = \sigma = \sqrt{n \cdot p \cdot (1-p)}.$$

In this case, the values are:

$$\text{mean} = \mu = 12 \cdot 0.24 = 2.88; \text{standard deviation} = \sigma = \sqrt{12 \cdot 0.24 \cdot 0.76} = 1.479.$$

The shape of the graph for the binomial distribution approximates the shape of the normal distribution when the number of individuals in the population is large. Although the binomial curve is slightly asymmetric, the graph for the normal distribution is symmetric.



Left: The graphs of three normal distributions with the same mean μ and standard deviations σ = 1; σ = 2; σ = 3. Right: graphs of three normal distributions with the same standard deviation σ = 1 and means μ_1, μ_2, μ_3, μ_4.

Statisticians and economists whose work involves applying probability theory must be familiar with a wide range of probability distributions. The study of each situation involving random variables (variables with values that cannot be predicted) corresponds to a determined probability distribution (a function).

There are some probability distributions that appear in a large number of economic and social problems. For cases in which the variable or character that is studied is discreet (integer values, such as yes or no) the binomial distribution can often be used. When the variable is continuous, the most suitable distribution, in many cases, is the normal distribution or 'Gaussian' bell curve.

A surprising mathematical result called the 'central limit theorem' states that if a population is large enough (e.g. the measurements of many axles manufactured by a machine, the voters in a country, the biometric parameters of human bodies, psychological characteristics and the behaviour of societies) the probability distribution will conform to the normal distribution, meaning that the means of the samples taken from the population will also be distributed in line with the normal law, with the same mean as the population. And the standard deviation of the samples extracted are also distributed in line with the normal distribution, with the same standard deviation as the population from which they are taken.

This theorem is the basis of the relationship between the results of a sample (few individuals) and the values of the larger and more complicated reality (population, many individuals) we wish to study, and indicates the fundamental importance of the normal probability distribution in statistical studies.

From the results of surveys periodically carried out on the active population, it is possible to infer values for the whole population that is being studied, with a certain margin of error, which has been previously defined and is referred to as the confidence level of the study.

The probability $P(z)$ that an event z occurs is calculated using the area bounded by the distribution function and the axis OX. Hence for the graph of

THE STANDARD NORMAL DISTRIBUTION

The standard normal distribution is the normal distribution simplified and standardised on the characteristic that is being studied. The scale of all the real values x_i of the variable that is being studied is shifted, subtracting the mean μ and dividing the result by the standard deviation σ. The standardised variable is referred to as z.

$$z = \frac{x - \mu}{\sigma}.$$

In this way, the set of values x_i centred on μ becomes centred on zero. The division by σ converts it into a standard normalised distribution, with standard deviation $\sigma = 1$. Hence, the same distribution (the same graph, the same function and the same table of values) can be used for any real distribution once the change has been applied to the variable, making the calculation of probabilities much easier. This change of variable is required to make it possible to use a single table for the normal law, which is used in all studies.

With this change of variable (from x to z) all the distributions of statistical variables that behave in line with the normal law are transformed into the same distribution function (the so-called probability density function) whose mathematical expression is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ with mean 0 and standard deviation 1.

this function, it is meaningless to talk of $P(0.45)$, because it is only possible to be sure of the values within an interval due to the errors of measurement inherent in any instrument and because the area is zero (a segment has area 0).

For continuous statistical variables, it is only meaningful to talk of the probability that a value z occurs within a determined interval $z \in [a,b]$ and then calculate the area using an integral of the density function, as follows:

$$PNT(a \le z \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz.$$

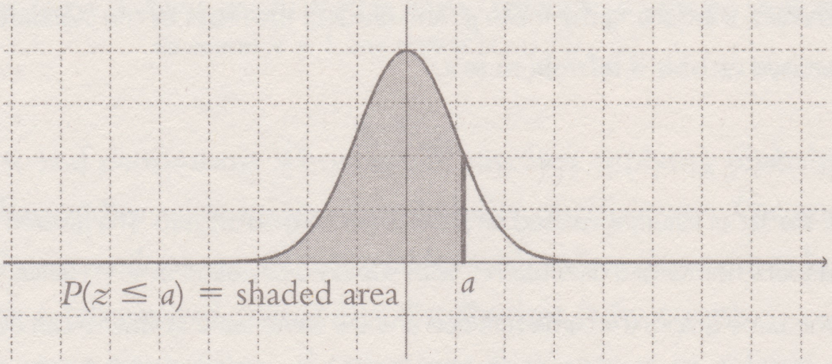
It is also possible to consider the probability that a measurement z is lower than a given value, or rather is an open interval such as $]-\infty, a]$, written as:

$$PNT(z \le a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz.$$

Both probabilities are calculated using an integral that gives the corresponding area.

Graphically, the probability that values of z greater than or equal to a occur is the shaded area of the following graph and corresponds to the value of the previous integral:

$$PNT(z \le a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz.$$



This integral must be calculated using mathematical approximations, since it does not have an exact value. These have been used to prepare the probability distribution tables for the standard normal distribution. They can be found in any basic book on statistics and can be used to solve any statistical study that follows a normal distribution, and hence to calculate probabilities such as:

$$PNT(z \le 1) = 0.8413;\ PNT(z \le 2) = 0.9972;\ PNT(z \le 3) = 0.9987.$$

Chapter 6

Mathematics and the Stock Exchange

The economy can be divided into two sections: one concerning the production of goods and services exchanged using money, which can be referred to as the 'real economy'; and another, different economy, which only involves the circulation of money and constitutes what is called the 'monetary economy', the world of finance. The financial markets are to be found within this realm of money, and there are various types depending on the sorts of assets traded: the foreign exchange market, the money market (interbank market), the market for derivative products, etc. The monetary economy includes stock exchanges (or capital markets), which are places where financial assets are bought and sold, including stocks, shares, bonds, public debt, etc.

Stock exchanges and stock markets

Stock markets are organised markets with rules of operation, and to which not all companies have access. Traditionally, transactions were organised in rings. At a certain time, the operators (exchange and stock agents) met on a specified 'parquet' to execute the transactions for each stock sector. In the same place, the brokers would communicate with those buying and selling securities, who gave them orders to buy and sell, and then carry out the transactions orally.

The transactions were carried out using an auction system according to the orders received. The number of shares for sale and the starting price were shouted out so that another operator with an order to buy from their client could also call out a quantity of shares and the price they were willing to pay. If they reached an agreement on the price, which meant the prices offered to buy and required for the sale were the same, the deal was closed, also orally. This was how the price of the stock was determined. The final prices often varied with respect to the starting price, since if there were more stocks than buyers, the value was lowered in order to find someone willing to buy them, whereas if the stock was oversubscribed, the prices would rise.

Two photographs of the New York Stock Exchange. The photograph on the left corresponds to the era in which all transactions were carried out orally, and the photograph on the right shows the current system, in which computer terminals prevail in stock exchanges.

Nowadays, the stock markets operate as a continuous market. The operators or financial intermediaries (the brokers) have a real-time electronic connection to stock exchanges.

There are various types of stock markets, which vary from country to country, since each stock exchange has its own regulations and is organised differently. The primary market, or issuing market, is the one on which newly created, recently issued securities are traded, including shares (variable profit) and bonds (fixed profit). The secondary market is that on which previously issued securities are traded, where investors come to resell their securities from a primary trade.

A large variety of products are traded on the financial markets. In addition to variable return securities (shares) and fixed-return private and public ones (e.g. liabilities, treasury bills, public debt), there is a wide range of financial products, including warrants, investment funds, options contracts, futures contracts... Financial products are released onto the market at a nominal value, and when they are traded on the market the price of the security varies depending on the demand.

Warrants are rights to acquire shares at a specified price and within a specific period. They can be warrants to purchase (*call*) that allow the holder to purchase a certain number of options in the future, and warrants to sell (*put*), which represent the right to sell a security within a given period of time. In Europe, the right of the warrant can only be exercised at the end of the contract, whereas in America, the rights of a warrant can be exercised at any point in time.

Stock indices are used to measure changes in the prices at which securities are sold. The indices are prepared using the changes of the index for each stock, based on 100 (and sometimes with base 1,000 or above). For a stock V_i, if at time t it has price $P_t V_i$ and we compare it with the price at which it was traded at time 0, $P_0 V_i$, the variation in its price with respect to time 0 can be calculated using an index I_{it} with base 100 at time 0:

$$\text{Index for stock } V_i \text{ at time } t\text{: } I_{it} = \frac{P_t V_i}{P_0 V_i} \cdot 100.$$

This would be the index for a single security. For a set of stocks, it is possible to calculate the simple average of the prices of the stocks at time t with respect to the simple average of their prices at time zero, which is the reference point for the base of the index. However, the most common method is to calculate weighted averages, or rather the index of each stock multiplied by its relative weight in the set of stocks that makes up the index. Various criteria can be used for the weighting, such as market capitalisation (the value of the company's capital or the value of the capital when it was first traded on the stock exchange) or trading volume.

In this case, the general index for a set of n securities will be the weighted sum of the indices for each security, applying a weighting criteria M for each stock i:

$$\text{General index} = I = \sum_{i=1}^{n} I_i M_i.$$

Various methods are used for calculating the indices, the most commonly used at present being the Laspeyres index, although some markets use the Paasche index (such as Paris or Brussels) or the Fisher index. Indices based on simple averages are now hardly ever used. Each stock exchange prepares its own indices in a different way, applying different formulae. Hence the New York exchange calculates the Dow Jones and the Nasdaq; London, the FTSE 100 (Financial Times-SE 100) or 'footsie' for short; Tokyo, the Nikkei; Frankfurt, the DAX; Paris, the CAC 40; Rome, the MIB 30; Brussels the BEL 20; and Madrid, the Ibex 35. The indices are calculated based on the changes in the weighted trading values according to the capitalisation or trading volume, which depend on the market in question.

They are often adjusted when one of the stocks of which they are composed is affected by a dividend payment or and issue of share rights in order to increase capital. Hence, for example, if a stock V_i is affected by the payment of a dividend d, the value of the dividend is subtracted from the price of the stock V_i, such that the adjusted index will be:

$$\text{Adjusted index for stock } V_i \text{ at time } t\text{: } I_{it} = \frac{P_t V_i - d}{P_0 V_i} \cdot 100.$$

There is a large variety of securities. They can be private (shares, bonds, shares in investment funds, IOUs from companies and lenders), or public (e.g. public debt, treasury bills and state bonds), fixed return (with a fixed interest rate) or variable return (which pay a variable dividend in line with profits obtained and their distribution), warrants, etc.

There is also a large range of transactions that can be carried out on the stock exchange. They can be spot transactions (settled at the time of the transaction), on credit (both purchases on credit using money lent and sales on credit with loan securities), term (settled on a specific date) or options – the buyer and seller undertake to buy or sell financial assets at certain advantageous conditions (with a premium), referred to as *call* (buy) *put* (sell) options respectively.

The underlying assets are the financial assets associated with an option. They can be stock indices, shares, commodities (e.g. gold, oil and raw materials), interest rates, foreign currencies, etc. Block trades are those that are carried out outside of a market. A buyer and seller agree to swap a certain number of securities at a specified price unrelated to the trading price. The quantity is often considerably large and is not traded on the market since this can cause high levels of volatility.

There are also futures contracts in which the two parties undertake to buy or sell certain financial assets at a specified price on a future expiry date. These are different from options contracts since in the latter case there is no obligation to buy or sell when the conditions of the contract are fulfilled, and the right can only be executed voluntarily during the period for which the contract is in force. However, for futures contracts the clauses of the contract are mandatory upon reaching the established expiry date. Futures contracts aim to provide coverage when the seller wishes to fix the price of the underlying asset and the market is on a downward trend. In this way, they will not lose money, whereas the converse occurs if the trading price of the asset rises. The buyer is interested in fixing the price of the asset when it is on an upward trend, to ensure it will be cheaper for them in the future. If the price rises, the seller with the futures contract will see their profits limited, whereas the buyer will gain because they will be able to purchase the underlying asset at a lower price.

Two other types of stock transactions are takeovers and initial public offerings (IPOs). The first occurs when a company wishes to purchase or take over another, and offers to purchase the shares of the latter at a price in excess of their value, at terms that are highly advantageous to the seller. The second refers to the situation when a company is floated on the stock exchange and offers its shares and liabilities on the market.

Now all stock exchanges operate electronically and have computerised trading mechanisms and transaction clearing houses. They also have management and transaction computer programs to facilitate both the work of brokers and operators to control and settle the operations that are carried out. However, they still have trading hours and operating regulations, meaning that to carry out transactions on them or offer assets for trading, it is necessary to comply with certain regulations that vary between stock exchanges and countries.

Fundamental analysis

Stock market analysts carry out two types of studies on the stocks that are traded on the stock exchange. The first consists of analysing the financial structure of the company that is being studied, or rather its balance sheets and other financial statements. The second is a technical analysis, based on statistical studies, paying specific attention to the graphical evolution of the trading prices, relating them to other explanatory variables in order to be able to make future predictions about the behaviour of the market for the stocks being studied.

Fundamental analysis often examines the balance sheet and profit-and-loss statements for a number of financial years, ranging from three to five, as well as studying the content of annual reports and information on the market, trade cycle and economic and sociopolitical environment. Above all, it is necessary to have an in-depth knowledge of the financial structure of the company by analysing its balance sheets and the relationships of its equity.

The balance sheet explains how a company is financed, how it obtains funds (capital, reserves, long-term credit, etc) and how it applies them. In principle, capital, reserves and long-term credit constitute what are referred to as permanent sources, which are applied to the investment in fixed assets (tangible and intangible) although when starting out, part of the permanent sources can also be applied to the acquisition of stock.

Fundamental analysis consists of relating the different collections of assets and liabilities on the balance sheet, both between each other and to profit and other off-balance sheet variables, in order to determine such things as solvency, liquidity, cash

THE STRUCTURE OF A BALANCE SHEET

The balance sheet of a company is organised in a table divided into two parts. On the left appear the assets, which list the goods and rights of a company in ascending order of liquidity. It is clear that the most liquid asset base is money available as cash or in a bank account; slightly less liquid are invoices pending payment by customers, all the way down to asset bases represented by land and buildings – the least liquid. The assets are divided into two parts. Fixed assets are made up of asset bases not subject to trading activity. Current assets are used for trading activities and their liquidity is less than a year.

The liabilities are listed to the right of the balance sheet and are placed in ascending order of the date they are due. The liabilities due first are those to short-term suppliers and creditors, which represent current liabilities, or debts recoverable within the year. These are followed by medium- and long-term debts (where due dates are greater than one year), reserves (profits that were not distributed during previous years) and, finally, the company's capital, owned by the shareholders.

	Assets	**Liabilities**	
Assets	**Fixed Assets**	**Not due**	**Permanent Sources**
Material	Land	Capital	
	Buildings	Reserves	
	Facilities		
	Machinery	**Long-term**	
Intangible	Patents, programmes	Debt	
Current	**Current assets**	Long-term credit	
	Stock	**Short-term**	**Current liabilities**
	Debtors	Short-term credit	
	Customers	Suppliers	
Available	Cash and current accounts		

The structure of a balance sheet.

flow, debt ratio and profitability. The ratios of the balance sheet represent the most important part of fundamental analysis.

One of the things that fundamental analysis reveals is that the stock of a company on the stock market (where market capitalisation = number of shares issued × price) is often much higher than the book value that is calculated for the balance sheet. The theoretical (or book) value of a share or stock in line with its balance sheet is:

$$\text{Theoretical share value} = \frac{\text{Net equity (capital + reserves)}}{\text{Number of shares}}.$$

Comparing this theoretical value with its price on the stock exchange makes it possible to determine if the company is overvalued or undervalued. In terms of the total value of the company, it is also possible to compare the value of its market capitalisation with the book value of the company:

$$\text{Price / book value ratio} = \frac{\text{Market capitalisation}}{\text{Book value}}.$$

Similarly for the cash flow or generated sources (profit + amortisations):

$$\text{Price / cash flow ratio} = \frac{\text{Market capitalisation}}{\text{Cash flow}}.$$

$$\text{Price / cash flow per share ratio} = \frac{\text{Price per share}}{\text{Cash flow per share}}.$$

FINANCIAL RATIOS

The solvency of a company is measured in terms of its working capital, which is the difference between current assets and short-term liabilities. It must be positive and as large as possible, although an excessive level of working capital represents the inefficient use of resources. If it is zero, this means the company has serious problems in terms of solvency. If it is negative, it runs the risk of having to suspend payments.

$$\text{Working capital} = \text{Current assets} - \text{Current liabilities.}$$

It can also be expressed as the complementary asset and liability bases: Working capital = Permanent sources − Fixed assets.

$$\text{Liquidity ratio} = \frac{\text{Current assets}}{\text{Short-term liability}} > 1.$$

$$\text{Liquid asset ratio} = \frac{\text{Current + available}}{\text{Short-term liability}} = 1.$$

If it is < 1 this indicates the risk of suspending payments. If it is much higher than 1, this indicates there is an excess of liquid funds and a risk of low profitability.

$$\text{Availability ratio} = \frac{\text{Available}}{\text{Current liabilities}}.$$

If it is small, a business may run the risk of liquidity problems.

$$\text{Debt ratio} = \frac{\text{Long- and short-term liabilities}}{\text{Liabilities}}$$

A suitable value is between 0.5 and 0.7.

$$\text{Ratio of financial expenditure} = \frac{\text{Financial expenditure}}{\text{Sales}} < 0.04.$$

Above 0.05 is regarded as excessive.

$$\text{Sales margin} = \frac{\text{Sales – Cost of sales}}{\text{Sales}}.$$

Indicates the trading margin of a company.

$$\text{Return} = \frac{\text{Profit before interest and tax}}{\text{Total assets}} > \text{Value of money.}$$

$$\text{Profitability} = \frac{\text{Profit}}{\text{Capital}} > \text{Value of money.}$$

$$\text{Payment term} = \frac{\text{Clients}}{\text{Sales}} \cdot 365.$$

Indicates the number of days it take to receive payments from customers.

$$\text{Payment term} = \frac{\text{Suppliers}}{\text{Purchases}} \cdot 365.$$

Indicates the payment term for suppliers and creditors in general.

$$\text{Asset rotation ratio} = \frac{\text{Sales}}{\text{Assets}}.$$

Indicates the volume of sales a company can generate in proportion to its assets. If it is >1, this indicates the number of times sales multiply assets.

$$\%\text{ increase in sales} = \left(\frac{\text{Sales } V_n}{\text{Sales } V_{n-1}} - 1\right) \cdot 100 \text{ between financial years } n-1 \text{ and } n.$$

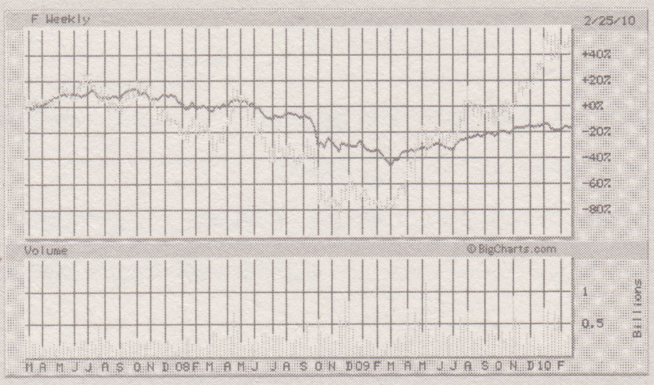
Technical analysis

Technical analysis consists of studying the specific behaviour of changes in the prices and values of various classes of financial assets, such as stock securities, indices, commodity prices, options, futures and foreign currencies by means of statistical and graphical analyses in order to predict future changes in the market for each of these elements. To do so, it is necessary to study the changes of the price variable of the securities over time using graphs (candlestick charts, bar charts, line graphs, point and figure charts, etc.). It is also important to carry out curve fitting, using statistical regression techniques that make it possible to relate the variable of price to other variables and hence make predictions with a certain degree of reliability. Technical analysis is based on three principles: history repeats itself; prices change as a result of trends; and movements of the market discount everything (i.e., prices

DOW THEORY

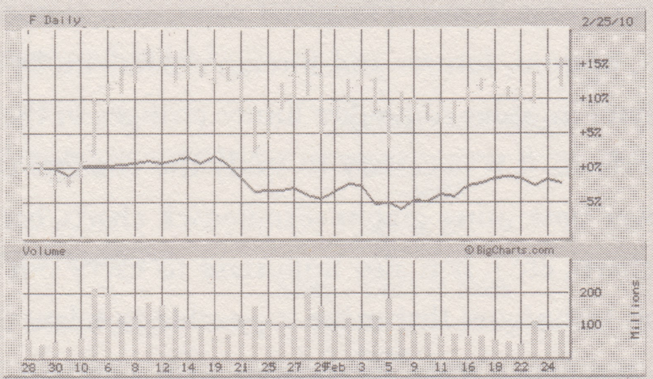
The first to analyse the oscillations of the market, paving the way for technical analysis, was the North American Charles Henry Dow (1851–1902), who devised the first stock index in 1884. To this day, his name is still associated with the famous stock index, the Dow Jones. Based on the graphical analysis of the behaviour of the indices, he developed his own theory on the behaviour of the market, distinguishing three types of trends: primary, secondary and tertiary.

The first is a change lasting for more than one year (between 1 and 3), although this can vary under certain circumstances, and can be upward or downward. After a downward period, a group of investors with initiative decide to invest in the stock market. As the market rises, other investors arrive *en masse* while the first group takes advantage of this to sell their stocks at a profit. After a while, as they begin to realise the increase is starting to fade, investors will progressively stop investing and the development will tend towards stagnation until a new downward period begins.



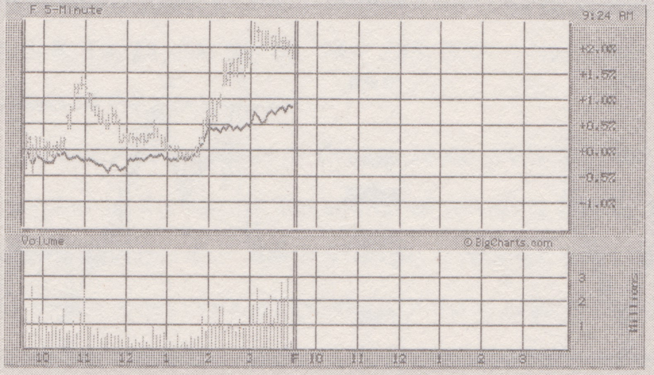
Primary Dow curve (over 3 years).

Secondary trends can last between one and eight weeks. They are shorter cycles that are superimposed on primary trends. They translate into small falls during upward periods and rises during periods when prices are falling.



Secondary Dow curve (over 2 months).

Tertiary trends have much shorter cycles and are the daily oscillations caused by the flow of supply and demand at a given moment in time.

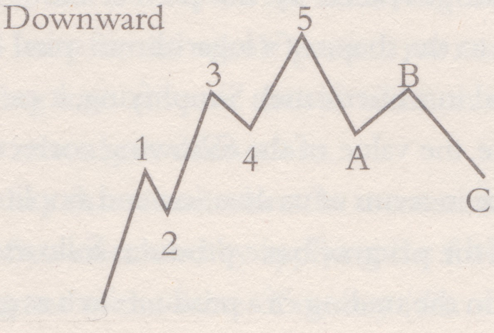


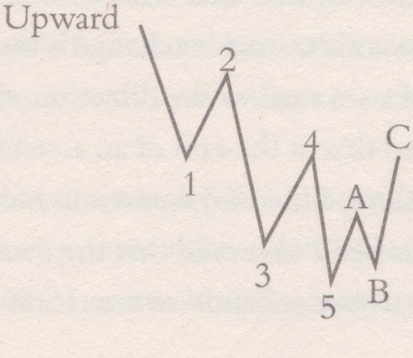
Tertiary Dow curve (over 2 days).

have already taken into account all information about the evolution of the company, microeconomic data, etc.).

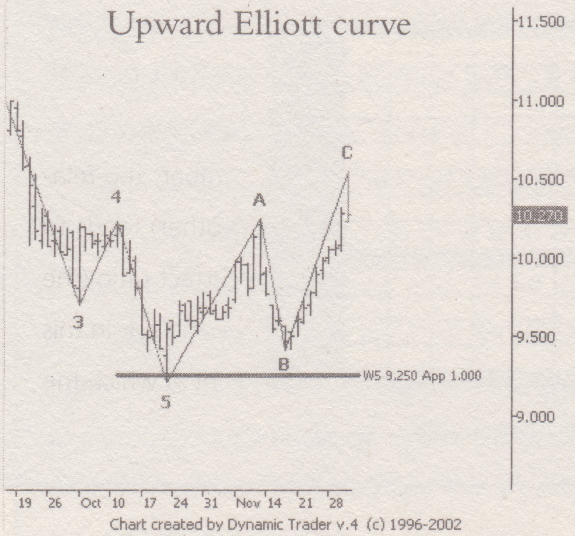
Ralph Nelson Elliott (1871–1948) made a subsequent contribution to Dow theory on the behaviour of the stock exchange. He observed that prices followed repetitive patterns that could be identified. The pattern was closely related to the Fibonacci sequence, which is a sequence in which each element is the sum of the two preceding elements. In 1946, based on a study of data over many years of the stocks of the Dow Jones Industrials Average (DJIA), he plotted a graph of the changes.

He concluded that the repetitive pattern had different periods and amplitudes. The Elliott Wave Principle states that the stock markets have a rhythm that is repeated: the rise has five upward waves, followed by three downward ones. The complete cycle has eight waves, five going up and three going down. In the ascending part of the cycle, the waves 1, 3 and 5 are impulsive (ascending), whereas 2 and 4 correct waves 1 and 3. After 5 ascending waves, a period of descent begins with three waves A, B and C. For the downward trend, the stock exchange also exhibits similar curves with five cycles.



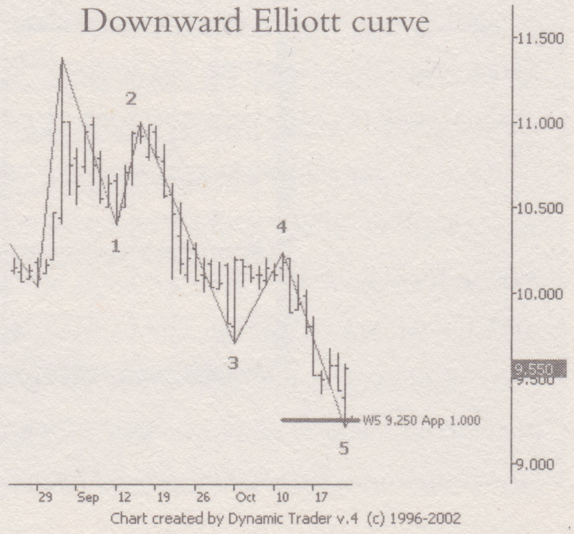


Elliott waves appear in real-world examples of the evolution of share prices, as can be seen in the graphs below.



Upward Elliott curve

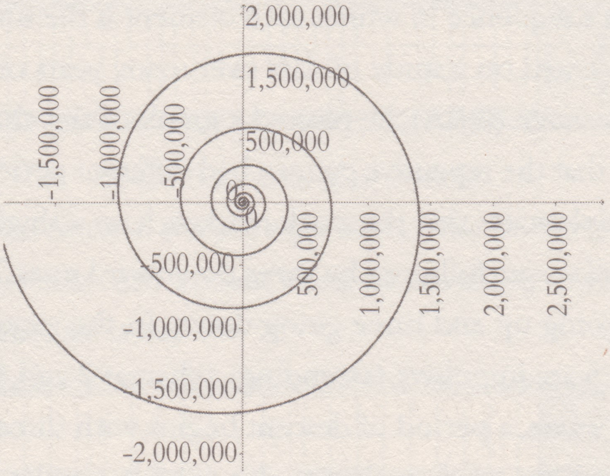
Chart created by Dynamic Trader v.4 (c) 1996-2002



Downward Elliott curve

Chart created by Dynamic Trader v.4 (c) 1996-2002

If we regard Elliott waves as a representation of human psychology producing the movements of the stock exchange on a large scale, it makes sense that the numbers of the Fibonacci series are precisely those that establish the scale of these waves.



Elliott claimed that the markets were also governed by the perfect ratio. The evolution of the stock exchange is thus similar to the shape of a logarithmic spiral (the shape of a sea snail or the direction of the wind in a hurricane). Simplifying, it can be predicted that at the end of an ascending phase, the value of the following correction will be three-fifths of the previous increase, both in terms of its duration and amplitude.

It has been observed that the evolution of the price of basic products follows the same structure as Elliot's waves. It can be seen in the trading of a product such as gold.

THE FIBONACCI SEQUENCE AND THE STOCK EXCHANGE

By studying the evolution of data for many stock exchanges, Elliott deduced that the mathematical basis for the wave principle is the Fibonacci sequence (discovered in the 13th century!). This sequence (1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377…) has many properties that can be applied to the analysis of the stock market.

The sum of any two consecutive numbers in the sequence gives the next number; the relationship of a given number to the next in the sequence (dividing one by the other) tends to approximate the irrational number $\Phi - 1 = 0.618034\ldots$, referred to as the perfect ratio (the golden ratio minus 1), which arises in the proportions of many objects in nature and art. In this way, a correction to a market trend that reaches a downturn of 61.8% is the point at which the market often begins to rise again; otherwise, the downturn will continue until 100%.

The square roots that do not give integers or rational numbers create a new class of numbers, referred to as irrational numbers due to the fact they cannot be written as a fraction. These numbers have an infinite number of decimal places with figures that have no cycle of repetition, referred to as periods, unlike a rational (although still infinitely long) number such as:

$$4/7 = 0.571428571428571428571\ldots$$

There are other types of irrational numbers, such as the number π, used in geometry, the number e, used to calculate compound interest, and the golden ratio Φ itself.

In recent times, its price has always moved against the trend of the stock market. As such, it must also follow Elliott's cycle, albeit in a different direction.

Measuring the performance of the stock market

At an IPO (initial public offering), there is an initial investment and periodical interest that is received for each year of the issue. To calculate the total profitability of the issue for the whole period, we must calculate the internal rate of return (IRR), taking into account that in addition to revenue, it is also necessary to account for the related tax deductions.

In a share issue, when the interest rate is variable, the overall profitability of the issue is often calculated by applying the formula for net present value (NPV) based on an estimate of the value of money (discount rate). The same method is used to value of a company, whereby the cash flows generated each year are updated in line with the initial investment and the discount rate that is applied.

One problem that often occurs is being able to know the annual profitability of an investment. Hence, for example, if at a time t an investment is made with a value I_0 and after n days the shares are sold for I_t such that $I_0 < I_t$ this gives a net yield of B from the sum of the surplus that is obtained $|I_t - I_0| > 0$ in addition to dividends, share rights, etc. The return R (in %) of the investment will be:

$$R\ (\text{Return}) = \frac{B}{I_0} \cdot 100.$$

The annual return R_a is calculated using a simple rule of three:

Return R obtained in ——————————— n days
Annual return R_a obtained in ——————————— 365 days

$$R_a = R \cdot \frac{365}{n}.$$

One of the most commonly used ratios in determining the return of a security is the PER (Price–Earnings Ratio), which relates the purchase value C_t of a security at a given time with the earnings obtained or expected to be obtained for the security over a period of time t (dividends, premiums, share rights, expected surplus), such that:

$$\text{PER} = \frac{C_t}{G_t} = \frac{\text{Share price}}{\text{Net earnings (earnings per share)}},$$

where C_t is the cost of the security, and G_t the net earnings expected for the security for the period t.

Note that the inverse of the PER is the return per share:

$$\text{Return per share} = \frac{\text{Net earnings (earnings per share)}}{\text{Price of share}} = \frac{1}{\text{PER}}.$$

In terms of the ratios of earnings obtained per share, the most common are provided below:

$$\text{Return per share} = \frac{\text{Total dividend}}{\text{Market capitalisation}} = \frac{\text{Dividend per share}}{\text{Price per share}},$$

where:

$$\text{Dividend per share} = \frac{\text{Earnings after taxes}}{\text{Number of shares}},$$

provided the full profit is always distributed among the shareholders. Where only part of the profit is distributed in the form of a dividend, it is useful to know the ratio represented by its percentage.

The dividend can also be calculated using the book value of the share, its market value or its nominal value.

Another ratio of interest is that which relates the cash flow of the business (sources generated = net profit + amortisations) to the number of shares:

$$\text{Cash flow per share} = \frac{\text{Net profit} + \text{amortisations}}{\text{Number of shares}}.$$

Finally, it is also necessary to mention how we measure the volatility and risk of securities. Risk is measured using a beta indicator (β), that shows the risk of investing in a given security and measures the variation (volatility) with which its price behaves with respect to the variation of the other prices on the market. The value n of β means that the return of the volatile security is n times the value of the return for the index or the market as a whole. Hence, if $\beta = 2$ and the average return of the market is 10%, the return for the security in question will be double, 20%.

Risk is measured by comparing the return of a security with the market as a whole. The return will exhibit certain variations with respect to the market average, measured using the standard deviation σ or the variance, which is the standard deviation squared, σ^2. If the mean return for the market is R, and the return for security i is R_i, the risk measured using the standard deviation will be:

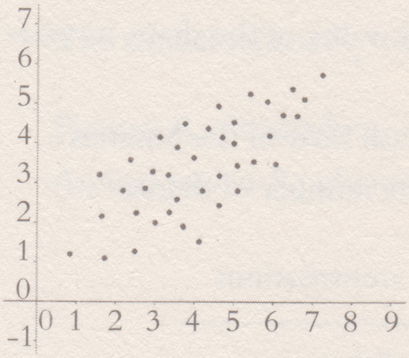
$$\sigma = \sqrt{\frac{1}{n}\sum(R_i - R)^2}.$$

It is also possible to use the semivariance, or rather the variance only taking into account the squares of the negative deviations, which measure the risk of losses.

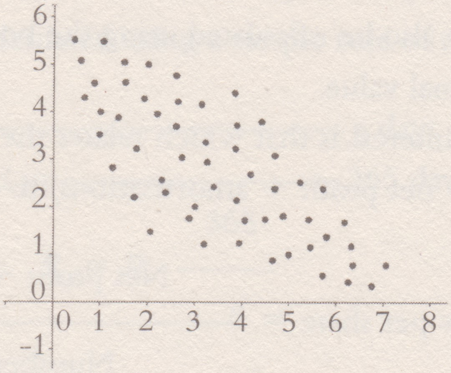
Correlation and regression

In economics, it is often the case that there are pairs of variables that appear related to each other. This joint variation is referred to as covariance. Essentially, we often say there is a covariance between two statistical variables when there is a unilateral causal relationship (X influences Y, but not the other way around) or interdependence (influence in both directions X on Y and Y on X).

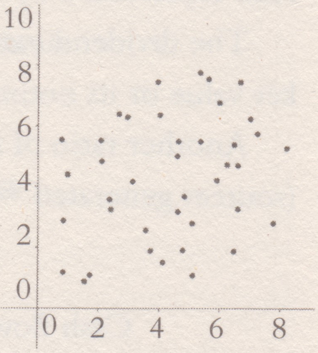
An initial study of the covariance between two statistical variables involves making use of graphical methods. The cloud of points that corresponds to all the pairs of values of the two variables under analysis has been plotted on the following graphs. This is called a scatter graph.



Positive linear correlation.



Negative linear correlation.



No correlation.

There are two techniques for analysing the covariance between two statistical variables: regression and correlation. Correlation calculates a coefficient (number) used as an indicator of the intensity or degree of covariance between the two variables, and regression determines a mathematical function that continuously expresses the covariance.

The statistical inference of the value of the correlation coefficient r for the population originating from a sample R is a problem that has been analysed in some depth. First of all, we can consider r as an estimate of R and then investigate if it is a good estimate. Given that it is possible to take many different samples from a population, each will have its own correlation coefficient. The r for all the samples constitutes a random variable with its own sample distribution.

It is also necessary to distinguish between correlation and causality. In cases such as the study of three variables (e.g. the number of books published, the number of televisions sold and the number of murders committed per year) it is possible to arrive at highly erroneous conclusions (e.g. the more books are sold, the more televisions will be sold or more murders committed).

CORRELATION COEFFICIENT

When we can predict that two variables are related by a linear correlation (which can be seen by drawing their scatter graph), we can then calculate the correlation coefficient. For a sample of *n* pairs of values (x_i, y_i) for $i = 1, 2, 3\ldots, n$, a measurement of the degree of linear association of both variables is given by the parameter *r*, called the linear correlation coefficient.

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sigma_x \cdot \sigma_y},$$

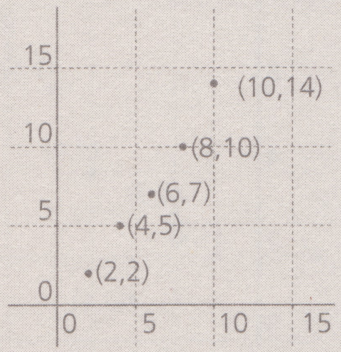
where $\bar{x}, \bar{y}, \sigma_x, \sigma_y$ are the means and standard deviations of the variables *X* and *Y* in the sample.

The correlation coefficient *r* always oscillates between −1 and +1. If $r = +1$ or $r = -1$, the points of the sample will be on a straight line. If *r* is a value close to + 1 (or −1) there is a strong linear relationship between the two variables. If *r* is small (close to 0) there is a weak correlation between the two variables, except when the scatter graph suggests another type of nonlinear correlation.

The coefficient *r* is a number with no dimensions (it does not have units or depend on those used by the variables *X* and *Y*).

The following table gives the values of two statistical variables *X* and *Y* for a sample of five individuals.

x_i	2	4	6	8	10
y_i	2	5	7	10	14



Scatter graph.

To analyse their correlation, we calculate the mean $\bar{x}$, the standard deviation σ and the index *r*.

$$\bar{x} = \frac{2+4+6+8+10}{5} = 6, \qquad \sigma_x = \sqrt{\frac{(2-6)^2+(4-6)^2+(6-6)^2+(8-6)^2+(10-6)^2}{5}} = 2.828$$

$$\bar{y} = \frac{2+5+7+10+14}{5} = 7.6, \qquad \sigma_y = \sqrt{\frac{(2-7.6)^2+(5-7.6)^2+(7-7.6)^2+(10-7.6)^2+(14-7.6)^2}{5}} = 4.128,$$

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sigma_x \cdot \sigma_y} =$$

$$= \frac{\frac{1}{5}\cdot\left[(2-6)(2-7.6)+(4-6)(5-7.6)+(6-6)(7-7.6)+(8-6)(10-7.6)+(10-6)(14-7.6)\right]}{2.828\cdot 4{,}128} = 0.981$$

REGRESSION LINES

If the points (x_i, y_i) that appear on a scatter graph for a statistical study approximate a curve or a straight line, an extremely useful mathematical exercise involves finding the mathematical function that best approximates to the cloud of points. The graph of this function is the regression curve (or line) that can be used for forecasting.

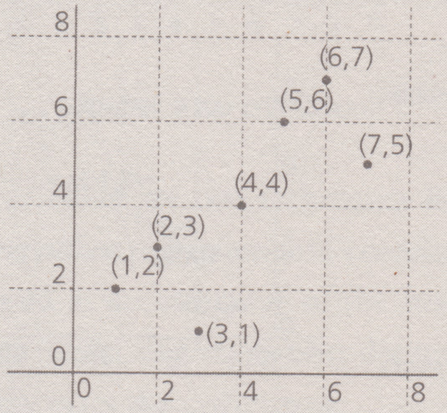
The mathematical method used to obtain the equation of the function that represents a cloud of points obtained from the values of the sample taken is referred to as the method of least squares.

$$\left.\begin{aligned} \sum y_i &= a\cdot\sum x_i + b\cdot n \\ \sum x_i y_i &= a\cdot\sum x_i^2 + b\cdot\sum x_i \end{aligned}\right\}$$

Solving this system of equations allows us to find the values of *a* and *b*. Let us assume, for example, they are, *a′* and *b′*. Hence, the line of best fit is the line of the equation: $y = a'x + b'$.

For the table and the diagram below, the regression line is calculated as follows:

x_i	1	2	3	4	5	6	7
y_i	2	3	1	4	6	7	5



$$\overline{x} = \frac{1+2+3+4+5+6+7}{7} = 4$$

$$\overline{y} = \frac{2+3+1+4+6+7+5}{7} = 4$$

	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
	1	2	1	4	2
	2	3	4	9	6
	3	1	9	1	3
	4	4	16	16	16
	5	6	25	36	30
	6	7	36	49	42
	7	5	49	25	35
Totals	28	28	140	140	134

$$\left.\begin{aligned} \sum y_i &= a\cdot\sum x_i + b\cdot n \\ \sum x_i y_i &= a\cdot\sum x_i^2 + b\cdot\sum x_i \end{aligned}\right\} \left.\begin{aligned} 28 &= a28 + b\cdot 7 \\ 134 &= a\cdot 140 + b\cdot 28 \end{aligned}\right\} \text{ Solutions: } a' = \frac{11}{14};\ b' = \frac{6}{7}.$$

The equation for the regression line *Y* over *X* is: $y=\frac{11}{14}x+\frac{6}{7}$. To calculate the regression line of *Y* over *X* we replace x_i by y_i to give the system (substituting the names of *a* and *b* to *c* and *d* to avoid confusion):

$$\left.\begin{array}{l}\sum x_i = c\cdot\sum y_i + d\cdot n \\ \sum x_i y_i = c\cdot\sum y_i^2 + d\cdot\sum y_i\end{array}\right\} \left.\begin{array}{l}28 = c28 + d\cdot 7 \\ 134 = c\cdot 140 + d\cdot 28\end{array}\right\},$$

which gives the same equation: $y=\frac{11}{14}x+\frac{6}{7}$. In the majority of cases there will be a different equation. In this case *X* and *Y* are interdependent.

Chapter 7

Growth and Development

There are two ways to study the economy. The microeconomic approach refers to the definition of economic variables defined at the level of the individual, affecting both consumers and domestic economies, and companies. The macroeconomic approach studies these same variables at an aggregate level. Hence the study of the prices, production costs and sales for a company, individual and family wages, levels of consumption, the sales of a product at a given price, and consumer preferences all form parts of microeconomics. In contrast, the study of the aggregate behaviour of all companies in a sector, both in terms of production and the use of labour, or the general level of pricing and the behaviour of aggregate supply and demand, form parts of macroeconomics. Macroeconomic variables relate to the evaluation of a country's production and its growth, supply and demand, imports and exports, savings, investment, public expenditure, inflation, etc.

Gross domestic product

A country's gross domestic product (GDP) is the total value of the market prices for the goods and services produced during a specific period, normally a year. It is calculated as the sum of the values from the production of all economic sectors, and as such is equal to the sum of all the value added by each sector. Therefore one sector of the economy buys goods and services from other sectors (raw materials, intermediate products, energy, technology, labour), which represent the costs of production, and then carries out a series of manipulations and operations to obtain a product (or service) with a higher price. In other words value is added to the production costs. The sum of all the added value created throughout a year by all sectors of production for a country is, as we have mentioned, the annual GDP, and its growth is one of the most commonly used metrics for evaluating a country's economic progress.

GDP can be measured in terms of the prices for each year, current prices, also using constant prices, which refer to a base year, hence neutralising the effects of inflation. This makes it possible to compare production for each year in real terms. Hence, for example if the GDP from the dairy sector was £500 million in 2005 and £700 million in 2010, representing an annual increase of 6.96%, the prices correspond to nominal GDP at the current prices for each year. However, if the annual rate of inflation over the five-year period has been 3%, £700 million in 2010 is equivalent to £603.83 million at 2005 prices. The real annual increase has been 3.85% and not 6.96%.

One part of the total production of goods and services (GDP) of a country is destined towards consumption, both private and public. Another part of the goods, which is not for consumption (such as machinery and industrial facilities or infrastructure), represents investment goods; and the proportion not consumed within the country is sold to foreign countries and so represents exports abroad. Production is evaluated in terms of the prices of sales, which are set by the market and are normally higher than the costs of the factors of production, increasing their value by adding value. To summarise, the GDP of a country over the course of a year will be:

$$\mathrm{GDP} = C + G + I + Exp,$$

where C = Private consumption, G = Government spending, I = Investment and Exp = Exports.

GDP is obtained from national accounts. Subtracting the value of imports from GDP gives the gross national product (GNP) at market prices, which represents the value that has really been spent in a country: its production less the difference between imports and exports:

$$\mathrm{GNP} = GDP - Imp = C + G + I + Exp - Imp,$$

where Imp = Imports.

Subtracting the depreciation in investments from GNP gives the net national product at market prices, and adding export subsidiaries and subtracting indirect taxes gives the net national product at the cost of the factors, also referred to as national income. Therefore,

$$\text{National income} = C + G + I + Exp - Imp - Depr + Sub - Ind\ tax,$$

where $Depr$ = Depreciations in investment, Sub = Export subsidies and $Ind\ tax$ = Indirect taxes.

Adding transfers, pension costs and subsidies paid to families, and subtracting social security payments, undistributed benefits and corporation tax gives personal income. Finally, subtracting direct taxes (income and inheritance tax) gives personal disposable income, which is the real measure of the standard of living and purchasing power of the citizens of a country.

Consequently, when we speak of national income or GDP, we must bear in mind that the two are not synonymous and, as we have seen, are extremely different large-scale accounting concepts. However, GDP, and specifically its rate of growth, is the most commonly used political indicator to express the health of an economy, both for developed countries and emerging or developing ones.

Per capita GDP (GDP/population) is also a widely used indicator, although it is also referred to as 'per capita income', despite really referring to GDP. The rate of growth does not always reflect good economic circumstances. Sometimes, excessive GDP growth, in the order of double digits, can entail 'costs of growth' that can result in serious structural imbalances in an economy, such as the incapacity of the financial sector, a loss of productivity, or a series of imbalances in transport and distribution infrastructure or services, which can curtail the expansive phase and even trigger a period of economic decline.

From a macroeconomic and national accounting perspective, a country's production can be considered in two ways: the value of production, as the sum of all added value generated by all the sectors that produce goods and services, or in terms of the total consumption of goods and services by the various agents in the economy. The total value of goods and services produced by a country is destined for consumption or the purchase of investment goods (machinery, infrastructure, equipment, etc.).

To summarise, the total value of production will be equal to the sum of all income (Y), and this will be destined for consumption (C) and investment (I), such that $Y = C + I$; and if we also bear in mind that the part of total income not destined for consumption constitutes savings (A), the previous equation for income can also be expressed as $Y = C + A$, or rather total revenue is equal to consumption less savings. Now, $I = A$; hence, in a closed economy (without taking into account foreign investment), savings will equal investment.

Keynes defined the function for consumption in terms of income, hence, $C = f(Y)$, and determined that at certain variations in levels of income there was a propensity to consume, defined as the quotient of the increases in consumption caused by increases in income:

$$0 \leq \text{Propensity to consume} = \frac{\Delta C}{\Delta Y} \leq 1,$$

such that it oscillates between the values 0 and 1, meaning it is always positive but less than one.

Keynes also defined the investment multiplier k that relates the increases in income to the increase resulting from investment:

$$\Delta Y = k \cdot \Delta I;$$

solving for k gives

$$k = \Delta Y / \Delta I.$$

Bearing in mind that $Y = C + I$, solving for I $(I = Y - C)$ and substituting the value of I into the equation for k, we have $k = \Delta Y/(\Delta Y - \Delta C)$. Dividing the numerator and denominator by ΔY gives:

$$k = \frac{1}{1 - \dfrac{\Delta C}{\Delta Y}},$$

The investment multiplier is equal to the inverse of one minus the propensity to consume. This means that the closer the propensity to consume is to one, the more sensitive consumption will be to increases in income, and the larger the multiplier. On the other hand, for high incomes, in which the propensity to consume decreases because consumers have already satisfied many of their needs and are no longer going to consume at the same proportion at which their income grows, the propensity to consume will be close to 0 and, as such, the multiplier k will fall.

Hence, for example, in a poor country the propensity to consume will be high, around 0.9, which means that 90% of income is destined for consumption. Applying the formula gives the multiplier k of 10. However, in a wealthy country the propensity to consume will be much lower, for example 0.6, meaning that only 60% of income is destined for consumption. In this case, the value of k will also be much lower, in this case 2.5. Hence, the investment multiplier is directly proportional to the propensity to consume.

The distribution of production by economic sectors

One of the most important factors to study when looking at a country's production is the distribution among the various sectors of production. Specifically, that means how the production of these sectors is interrelated, and what part of the production of each sector is destined for each of the various sectors as intermediary products and what part of the production is destined for final demand.

In this respect, the contribution made by Wassily W. Leontief (1906–1999) is significant. He, as an adherent of the *Tableau Économique* of the physiocrat François Quesnay, and basing his work on Walras' mathematical studies of general equilibrium, used a matrix to represent the inter-industrial relationships of a country's economy. In this, the rows represented the value of production (outputs) of a sector, which were then consumed by other sectors, and the columns represented the purchases (inputs) of each sector from the other industrial sectors. The inter-industrial relationships represented using this input and output matrix, referred to as input-output tables, were specified in the individual entries, such that the entry x_{ij} represents the total value of the production from sector i used by sector j, and X_i the total value of the production i sold to the other sectors of the economy.

Overleaf is an example of the structure of the input-output table for a country's production. The production levels of the various sectors have been ordered by rows (1, 2, 3… n), which include the agricultural, industrial and tertiary sectors. On the other hand, the columns represent the consumption of the various sectors.

		Sectors of Consumption (columns: inputs)									
		Intermediate consumption							Consumption		
		1	2	3	4	…	n	Production	Internal demand	Exports	Σ Final consumption
Sectors of production (rows: outputs)	1	x_{11}	x_{12}	x_{13}	x_{14}		x_{1n}	X_1			
	2	x_{21}	x_{22}	x_{23}	x_{24}		x_{2n}	X_2			
	3	x_{31}	x_{32}	x_{33}	x_{34}		x_{3n}	X_3			
	4	x_{41}	x_{42}	x_{43}	x_{44}		x_{4n}	X_4			
	n	x_{n1}	x_{n2}	x_{n3}	x_{n4}		x_{nn}	X_n			
Σ Intersectorial purchases		X_1	X_2	X_3	X_4		X_n	X			
Taxes on products											
Wages											
Surplus											
Total internal production											
Imports											
Σ Supply resources											

Input-output table. Table of intersectorial relationships.

The input-output table is a highly effective instrument for national accounting, estimating gross national product and forecasting. The total value of production can be obtained from the intermediary consumption matrix. Part of the production is destined for internal consumption (internal demand), or rather private consumption, in addition to government consumption (public expenditure) and investment (gross formation of capital). The surplus part not consumed by internal demand is destined for export. The total of internal demand and exports constitutes the final production or total output. From the point of view of consumption (columns), adding taxes on products (value added tax, etc.), taxes on wages and business earnings (profits) to intersectorial consumption gives total internal production, to which imports are added to give the total value of resources. An important use of input-output tables is the study of the level of technology in an economy through the definition of the technical coefficient matrix. Each entry where the rows and columns meet between sectors is allocated a technical coefficient for sector *i* with respect to sector *j* of the following form:

$$a_{ij} = x_{ij}/X_j.$$

The technical coefficient matrix is extremely useful in economic forecasting, since it makes it possible to evaluate the impact of a certain event on an economy, such as a war or the relocation of a sector of the economy.

Economic cycles

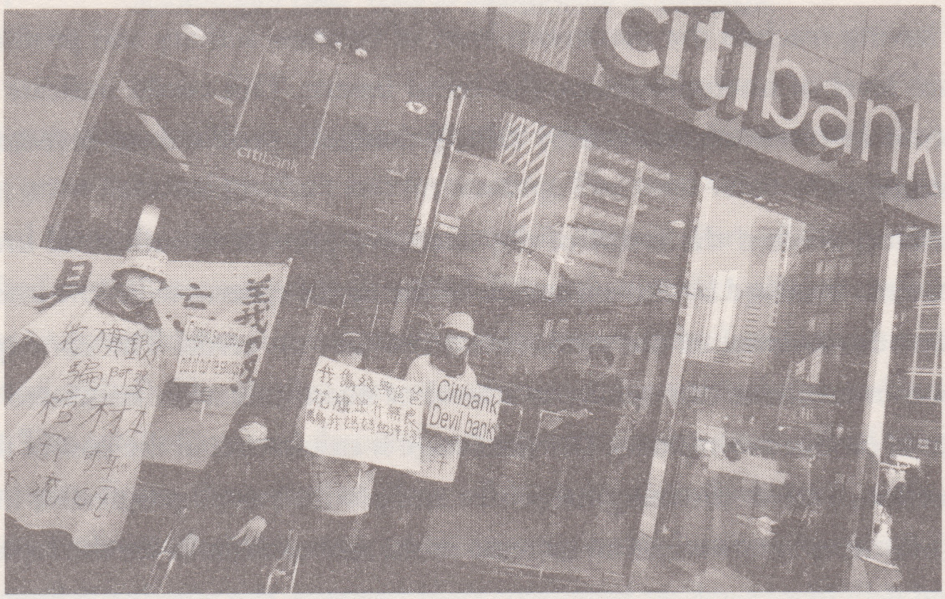
Economic cycles are the periodic phases of acceleration and deceleration of the economic activity of a country, economic region, group of countries, or even the world itself, in which growth in production, business profits and employment oscillates between the expansion and prosperity of the population and its companies, and recession and poverty. Economic cycles can be represented graphically by means of a sine wave with GDP on the vertical axis and time on the horizontal axis. The distance between the maxima and minima of the curve indicates the frequency and duration of each cyclical period.

The cycle begins after a period of crisis with a phase of recovery and subsequent expansion, until reaching a higher maximum or peak. After a certain period of time of having reached the maximum level of business activity and profits and the minimum level of unemployment, the trend inverts. This could indicate the start of a period of recession, caused by falling demand, production, investment and employment.

Recession is often the result of the excessive growth in industrial investment, which in turn is caused by the expansion of the credit system, with low interest rates that favour the renewal of the productive structures on credit. The consequence of this is over-investment that results in a level of overproduction associated with the underuse of productive capacity. The next stage is an increase in the cost of the means of production, a fall in the marginal profits and productivity, a progressive increase in the rate of unemployment and an increase in the rate of defaults for the repayment of credit and debt in general. As unemployment increases, the sales made by companies fall, and this reduces the purchasing power of demand, further increases unemployment, reduces inflation through the reduction of production, freezes or reduces salaries and the prices of products, and causes the profits and forecasts of businesses to fall.

The latter results can shake the financial markets and cause values to fall on the stock markets. The period of recession or depression becomes more acute, worsening until it provokes a period of crisis and reaches the minimum point of the cycle. Once this phase has been reached, the economy arrives at a point at which it can fall no more and which corresponds to the minimum capacity required to ensure the country continues to function. In this phase, the least competitive companies have already disappeared and, to put it somewhat bluntly, the economy has disposed of its least efficient elements. At this point, it is only a matter of time before, as the result of governmental measures to stimulate investment and employment, or rather the renewal of demand, the machinery of production is reactivated to enter a new expansive phase and repeat the cycle.

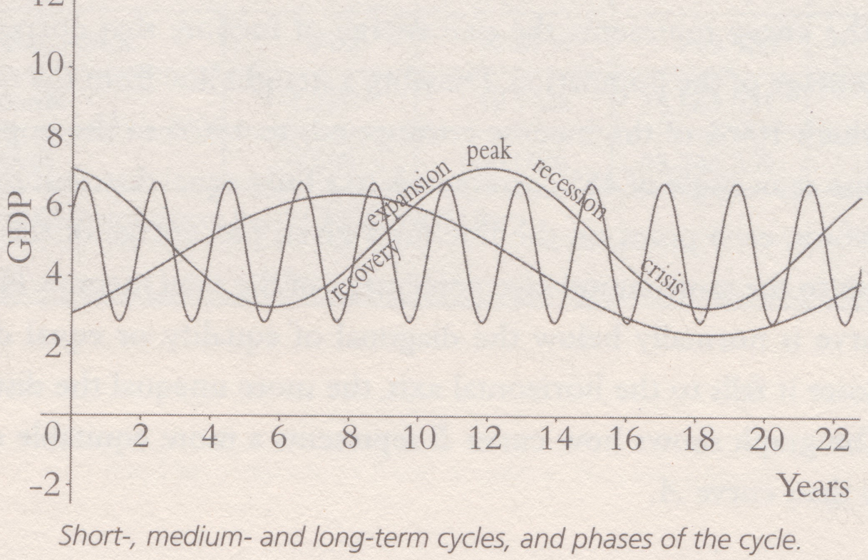
It should be pointed out that the type of crisis described represents the real economy, or rather the sectors that produce goods and services. However, the crisis can also originate in the financial sector, when speculative financial bubbles burst, as well as in certain fields of financial investment, as a result of a lack of control over the financial markets and an excess of investor euphoria. This can be caused by a certain climate created by the financial sector itself. This type of economic crisis, one with a financial origin, often has harmful consequences on the real economy, since it affects the solvency of banking institutions. Then lack of liquidity damages the normal operation of companies and consumer demand, and can delay the return to recovery.



A customer protest at a branch of Citibank in Hong Kong, owned by Citigroup, the largest financial group in the United States prior to the bankruptcy of various financial institutions related to the 2008 global crisis. The crisis was triggered by the failed mortgage market.

The foundations of theories concerning economic cycles lie in long-term statistical series. It has been shown that there are various types of economic cycles with different durations, which can overlap such that if the crisis points of the different classes of cycles coincide, they will become more acute. Numerous statistical studies have been carried out on the cycles, and attempts have been made to use regression and multivariate analyses to investigate which variables determine the causes of economic crises. Economic models have also been used frequently to predict future changes of trends in the economic cycle. These models consist of a set of equations involving certain variables (e.g. GDP, prices, exports, demand, demand for money) in line with the behaviour of a series of external variables. It is also necessary to consider crises caused by events that lie beyond the economy, such as the 1973 oil crisis or the invasion of Kuwait by Iraq in 1990.

Empirical studies have identified various types of economic cycles. In addition to seasonal cycles in some sectors, such as agriculture and tourism, there are essentially three types of cycles: short, medium and long term. The short-term cycles include those described by Kitchin; their duration varies between 18 and 40 months. Medium-term cycles vary between 4 and 5 years (as described by Mitchell), and between 6 and 10 years (described by the Frenchman Juglar). Finally, the Russian Kondratieff studied extremely long-term cycles, between 50 and 60 years, the period in which, according to the author, crises as severe as that of 1929 occur. As we saw in the previous chapter, cycles also occur on the stock exchange in the prices of stocks and indices. These were discovered by Charles Dow and subsequently by Elliott, who defined eight waves of stock market oscillation.



Short-, medium- and long-term cycles, and phases of the cycle.

Measuring the well-being of a country

There is a widespread belief that the well-being of a country is measured by its level of per capita income (actually GDP per capita), by the rate of GDP growth, or the surplus of its balance of trade (export of goods minus imports), the balance of its current account (adding the balance of trade to the balance of services made up of transport and freight, tourism, insurance, etc.) or its balance of capital (the entry and exit of capital). When it comes to the balance of payments, it should be borne in mind that the balance will always balance out, since a current account surplus will be offset by a deficit in the balance of capital, and vice versa.

One aspect closely related to the balance of payments is external debt. In developing countries and those with high poverty indices, in which the population grows at a rate in excess of GDP, it is common to see a higher level of national indebtedness, especially to wealthier powers.

On the other hand, countries with high incomes are limited to extremely restricted areas: the European Union, Switzerland, the Nordic countries and small European states, the United States and Canada, Japan, South Korea and Australia, in addition to the oil exporting countries of the Persian Gulf. Wealth is extremely poorly distributed throughout the rest of the world.

However, even within a country, wealth can be distributed in a highly unequal manner. One analytical instrument for measuring the distribution of income in a country is the curve designed by the mathematician Lorenz, in which a graph is used to represent the percentages of income that corresponded to certain percentages of the population. The vertical axis represents the cumulative percentages of income and the horizontal axis represents the cumulative population percentages. Each point on the curve represents the percentage of income that corresponds to a given percentage of the population. Drawing a straight line from the origin to the point at which 100% of the income corresponds to 100% of the population, the diagonal line at an angle of 45° corresponds to a fully equal distribution of wealth. In other words, each point on the line for a given percentage of the population corresponds to the same cumulative percentage of the total income. However, the Lorenz curve is normally below the diagonal of equality, or equal distribution, and the closer it falls to the horizontal axis, the more unequal the distribution of income. The graph shows how curve *B* represents a more equitable distribution of income than curve *A*.

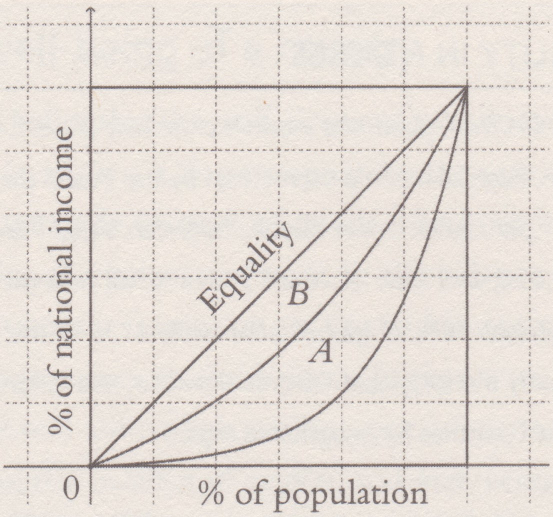
GLOBAL INEQUALITY IN FIGURES

Equality between countries throughout the world is extremely large. Countries are classified using categories based on levels of income established by the World Bank (high income: greater than or equal to $9,361 per capita; upper middle: between $3,031 and $9,360, lower middle: between $761 and $3,030; and low: less than $761). It has been observed that 16% of the world's population produces 76% of global GDP, while 37% of the world's population with low income produces only 3% of global GDP, as shown in the following table, which groups low- and middle-income countries by geographic region.

Global production and population 2006

Groups of countries	Population (millions)	% Global population	GDP ($ million)	% global GDP
Low income	2,403	36.87%	1,611,831	3.34%
Upper middle income	810	12.43%	5,316,864	11.02%
Lower middle income	2,276	34.92%	4,734,576	9.81%
Total middle income	3,086	47.35%	10,051,440	20.83%
Middle and low income	5,489	84.21%	11,663,271	24.17%
East Asia and Pacific	1,900	29.15%	3,636,593	7.54%
Europe and Central Asia	460	7.06%	2,483,602	5.15%
Latin America and Caribbean	556	8.53%	2,945,193	6.10%
Middle East and N. Africa	311	4.77%	730,103	1.51%
South Asia	1,493	22.91%	1,142,319	2.37%
Sub-Saharan Africa	770	11.81%	709,500	1.47%
High income	1,029	15.79%	36,583,031	75.83%
World total	6,518	100.00%	48,246,302	100.00%

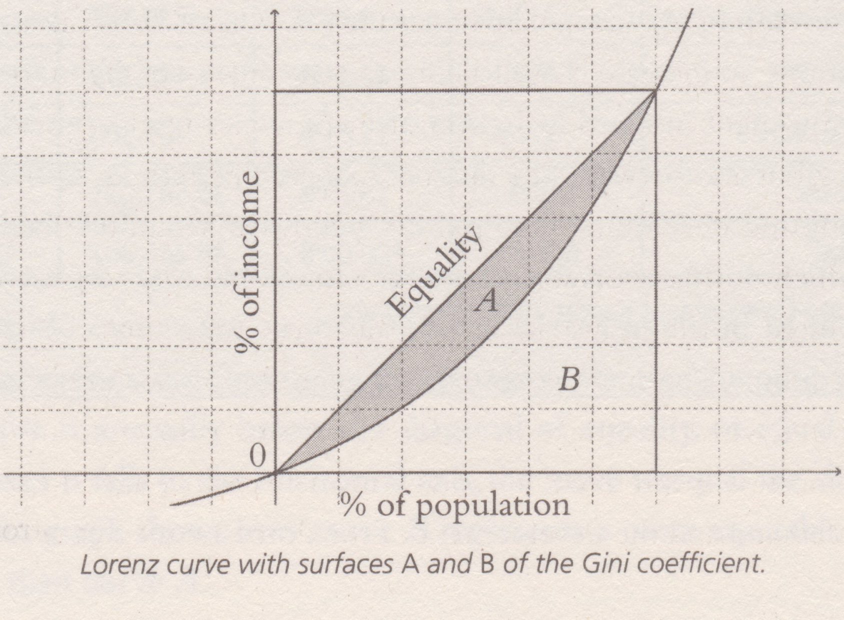
Source: World Development Report 2008, World Bank.



Lorenz curve.

The Italian mathematician Corrado Gini used the Lorenz curve to establish the coefficient that bears his name for measuring inequality. Once the Lorenz curve has been plotted, two areas are identified: area *A*, which corresponds to the area between the diagonal of equality and the curve, and area *B*, which corresponds to the area of the surface below the Lorenz curve down to the axes. The Gini coefficient is defined as the quotient of the surface of *A* and the total surface *A* + *B*, hence:

$$\text{Gini coefficient} = \frac{\text{Area between Lorenz curve and the line of equality}}{\text{Total area of triangle below the diagonal}}.$$



Lorenz curve with surfaces A and B of the Gini coefficient.

$$\text{Gini coefficient} = \frac{\text{Surface of } A}{\text{Surface of } A + B}.$$

The smaller the area of A, the more equal the distribution of income. So, as it tends to zero, it will grow ever closer to the diagonal of equality, and the value of the Gini coefficient will also tend to zero. On the other hand, the greater the area of A, the larger the Gini coefficient and this will tend to 1. The coefficient is used to measure inequality, not just in terms of the distribution of income, but also for regions and countries, as well as for measuring the inequality in the distribution of certain social services or public expenditure among the population of a country or region.

An important concept that must be taken into account in the development of a country is its level of economic development. What does it mean to say that one country is more developed than another? What variables should be used to classify countries as wealthy and poor? The rate of GDP growth is important, but less developed countries often exhibit much higher rates of growth, since they obviously start from a much lower gross domestic product, whereas developed countries generally exhibit low levels of growth.

The first question asked by international organisations and United Nations agencies in order to tackle poverty and the lack of economic development is what economic measures should be taken into account to ensure an underdeveloped (or developing) country can make progress on the path to economic development. Economic development is not solely limited to GDP growth, but also to improving social conditions, an increase in well-being and the quality of life of the population, and an improvement in the distribution of income and wealth, not only between social groups, but also between the different geographic regions of the country.

For economic development to exist, a primitive accumulation of capital is required (which normally originates in agriculture), in order to generate surpluses that are invested in the industrial sector, thus increasing the productivity of the country's economy. The increase in exports generates an accumulation of capital, which together with the attraction of foreign capital is invested in the industrial sector. The increase in productivity entails an increase in wages, in addition to an increase in demand and the development of the domestic market. Strong growth is accompanied by the appearance of demand for better quality services, which, in the final phase of development will trigger the creation of a significant tertiary sector, the productivity of which exceeds the industrial sector.

At a practical level, economists ask what are the variables that define economic growth and, in their most elaborate stage, the economic growth of a country. The most important variables are gross domestic product (GDP), capital stock, communications and transportation infrastructure (so-called 'social capital'), expenditure on education (human capital), the workforce, savings, the level of income of demand, the level of technology, the structure and size of markets, etc. However, the fundamental question is why do countries exhibit economic growth over the course of time? Many of the answers are based on investment, the level of technology, and improved education, which in turn results in a more qualified and productive workforce, or the size of the workforce itself, etc.

As a simplification of the problem, a series of equations have been defined that relate all these variables and use econometrics to measure and verify them using statistical techniques, such as linear regression and curve fitting. The econometric method consists of specifying a model using various equations that represent functions relating a series of explanatory and dependent variables, which must be evaluated and verified in order to make forecasts using initially estimated parameters (constants in the equation whose values must be empirically estimated).

A simple econometric model would be that of supply and demand in an active market. Let q_1 be the level of demand for a given product at a price p_1 and q_2 the level of supply of the product at the same price p, both q_1 and q_2 are functions of price $f(p)$ and $g(p)$ respectively and are expressed in this way: demand function: $q_1 = f(p)$, and supply function: $q_2 = g(p)$. The functions $q_1 = f(p)$ and $q_2 = g(p)$ can be expressed in the form: $q_1 = \alpha + \beta\, p$, $q_2 = \omega + \eta \cdot p$, where α, β, ω, η are the parameters to be estimated using empirical data. One of the conditions that must hold is that both supply q_2 and demand q_1 for the product in question must coincide, since they are the levels of demand and supply at the point of equilibrium $q_1 = q_2$. The regression analysis of the clouds of points that correspond to the statistical data observed in the variations of the levels of demand and supply for the product in line with the variations in price make it possible to estimate the values of the parameters and define equations that can be used to predict the levels of supply and demand for the product in line with the variation in their price.

Indicators of wealth, economic development and human development

Both the rate of growth of gross domestic product (GDP) and its per capita level are insufficient indicators for correctly measuring a country's level of economic development. This is especially true when taking into account the enormous structural differences and variations in the quality of life that exist between rich and poor countries. The United Nations Development Programme (UNDP) has defined and calculated various indicators that make it possible to gain more rigorous and in-depth knowledge of the real differences between the various levels of development between countries.

The UNDP has developed various indices, which include the Human Development Index (HDI) and the Human Poverty Index (HPI). The former is a composite statistical index that measures the progress of a country in terms of three basic dimensions of human development:

1. A long and healthy life, measured in terms of life expectancy at birth.
2. Access to knowledge, or education, measured by the literacy rate among adults and the gross combined rate of enrolment in primary, secondary and higher education, in addition to the duration of compulsory education.
3. A decent standard of living, measured by GDP per capita and purchasing power parity (PPP) measured in dollars.

The United Nations defines human development as the process by which a society improves the standard of living of its inhabitants by increasing the goods that allow them to meet their basic and supplementary needs, and the creation of an environment in which human rights are respected. Human development is the number of options available to someone to be who they wish or do what they want. The greater the number of options, the greater the level of human development; the fewer the options, the lower the level of human development. Human development can also be defined as a way of measuring the quality of life of people in the environment in which they live. It is a fundamental variable for classifying a country or region.

In the report published by the UNDP in 2011, the HDI varies between Norway, with an index of 0.943, in first position, and D.R.Congo, with an index of 0.286 in 187th position. The UNDP classifies countries into three main groups:

Countries with a high or very high level of development ($HDI \geq 0.700$): 94 countries. Countries with medium development ($0.500 \leq HDI < 0.699$): 47 countries. Countries with a low level of human development ($HDI < 0.500$): 46 countries.

The countries with the highest levels of human development include the developed countries of North America, Western Europe, the Nordic European countries, Switzerland, Japan, and some countries of the southern cone of Latin America; whereas the countries with low levels of human development include practically all those in sub-Saharan Africa.

HDI arose as part of an initiative to classify countries based on variables that differ from those normally used in economics (GDP, balance of trade, energy consumption, unemployment, etc.) and which corresponded to other areas of society that determine quality of life, such as education (literacy rate, number of enrolments by educational level, etc.) and health (birth rate, life expectancy, etc.). The value of the index is determined by the weighting of the following variables: life expectancy at birth, literacy rate among adults (percentage of people aged 15 and over), gross rate of enrolment in primary, secondary and tertiary education, and GDP per capita.

The UNDP also prepares other indices, such as the Human Poverty Index (HPI), calculated based on the probability of living for 40 years, the rate of illiteracy, children whose weight is below the average for their age, and the proportion of the population living on less than $2 a day. It also prepare indices on environmental factors (CO_2 emissions) and other areas (the consumption of conventional, renewable and nuclear energy, military expenditure, human rights, women's rights and gender equality, etc.).

The method for calculating HDI first requires the creation of an index for each of the variables it considers (life expectancy, education and GDP). To do so, minimum and maximum values (limit values) are selected for each indicator. Each of the components is expressed by a value between 0 and 1, using the following general formula.

$$\text{Index for component} = \frac{\text{Real value} - \text{Minimum value}}{\text{Maximum value} - \text{Minimum value}}.$$

HDI is calculated by taking the average of the three main components. The limit values (maximum and minimum) used for calculating HDI are 85 and 25 years for life expectancy at birth; 100% and 0% for the two components of education; and $40,000 and $100 for GDP in dollars per capita adjusted for each country for PPP with the US dollar.

Appendix

As has been shown, the study of economic growth and development uses equations, systems of equations, functions, algorithms, graphical representations, matrix calculations, etc. This book has shown the key mathematical concepts that are used in economic analysis. Throughout the history of the various schools of economic thought, mathematical analysis has been progressively used both to develop and prove various economic theories. With the passage of time, the use of the inductive method has been imposed and economic theories have had to be based on the analysis of real data that has been collected. This requires an initial calculation, logically facilitating, in a second stage, the use of statistical and mathematical methods for analysis.

Differential calculus and derivatives by the marginalist neoclassical schools, whose key figures included Léon Walras, Wilfredo Pareto, Stanley Jevons and Alfred Marshall, was used in developing the theories of consumer behaviour and the formulation of general equilibrium. Meanwhile, systems of equations to define the economic models created econometrics, a discipline that arose from the combination of economic analysis and mathematical and statistical analysis, or rather which applies these types of analysis to microeconomic theory to check its compliance. Finally, matrix calculations were used by Leontief in developing input-output tables to study the intersectorial relationships of the economy.

To summarise, the past century has witnessed the progressive use of mathematics in economic analysis, led in many cases by the interest of mathematicians themselves in studying the science of economics. In this respect, we must remember the contribution made by great mathematicians to the advance of the science of economics, such as John Maynard Keynes, the founder of macroeconomics who inspired the development of economic models, and John von Neumann and Oskar Morgenstern, who developed game theory. The development of economics as a science is intrinsically linked to the increasingly rigorous use of mathematical methods. Hence the reason why many departments of economic sciences and business schools require a solid mathematical grounding for entry to the disciplines.

Bibliography

Halm, G.N., *The Economics of Money and Banking*, Homewood, R.D. Irwin, 1961.

Ifrah, G., *The Universal History of Numbers,* London, Harvill Press, 2000.

—, *The Universal History of Computing*, Hoboken, John Wiley & Sons Inc., 2001.

Krugman, P.R. and Obstfeld, M., *International Economics. Theory and Policy.* London, Pearson Educational, 2010.

Samuelson, P.A. and Nordhaus, W.D., *Economics*, New York, McGraw Hill, 2009.

Stigler, S.M., *The History of Statistics*, Cambridge, Harvard University Press, 1989.

Sydsaeter, K. and Hammond, P.J., *Essential Mathematics for Economic Analysis*, London, Pearson, 2008.

Index

accounting 8, 22-23, 138-139, 142
analysis
    fundamental 121-125
    technical 121, 125-129
arithmetic 15, 17-20, 25-30, 38

calculating algorithms 8, 18-19
coefficient
    correlation 132-133
    Gini 148-149
    technical 142-143
convertibility 32-34
correlation and regression 132-135
cost
    marginal -77, 101
    average 74-77
curve
    demand 46, 98-99, 107
    Lorenz 146, 148
    Phillips 46, 47
    supply 87, 88, 99
cycles
    economics 9, 143-145
    isoquant 73
    Juglar, Kitchin, Mitchell, Kondratieff 146
    costs (average, marginal) 69-78
    production 73
    supply and demand of labour 87-88
    wage 87

derivatives 76, 77, 101
double entry 7, 21
Dow, C.H. 126-127, 145
Dow theory 129
    interbank interest 50

equations and inequations 40
elasticity of demand 105-107
Elliott, principle of waves 127, 129

factors of production 69, 72, 73, 77-79, 100, 102
Fibonacci (Leonardo of Pisa) 18, 40, 61, 127-129
financial ratios 124
fraction 11, 18-20, 42, 129
function
    cost 74
    production 69-73
    linear 71
functions 42, 53, 69, 70-72, 150

Galbraith, J.K. 86
GDP per capita 139, 146, 152, 153

hyperinflation 47-48

iron law of wages 84
index
    Dow-Jones 126
    Fisher 45, 119

  general stock 120
  Laspeyres 45, 119
  Paasche 45, 119
  price 44

inflation
  galloping 48
  rate of 45, 56, 80-82, 137, 138
input-output tables 24, 141-143
instalment of a loan 58-61
insurance premium 66-67
invoice discounting 109

Keynes, J.M. 26, 84, 140

Lavoisier, A. 27
Leibnitz' calculator 28
Leontief, W. 141
lines and curves of regression 125, 134-135
logarithm 58, 64, 154

Malthus, T. 28, 83, 84
market equilibrium 97, 100-105
Marx, K. 25, 84
Menger, C. 25, 102
market
  capital 49
  foreign exchange 34
  labour (supply and demand) 83-86
  stock 111
monetary standard 7, 32-34
multiplier 140-141

national income 137-141
numbers
  Greek 14
  negative (debt or absurd numbers) 12, 18, 21-23

optimal production level 69-78

Pacioli, L. 21
Pareto, V. 102- 105
purchasing power parity 36, 152
Pascal, B. 7, 26
Pascal's calculator 7, 26, 27
purchasing power 44, 56, 80, 81, 86, 87
Price Earnings Ratio (PER) 130
Gross Domestic Product (GDP) 137-141, 143, 145-147, 149-154
product
  derivatives 79, 117
  geometric 59-61
  marginal 78, 86-87
profit
  marginal 102, 143
  residual (Residual Income) 80

rate
  annual equivalent interest rate (AER) 56
  exchange (floating or fixed) 34-36
  internal rate of return (IRR) 83, 129
  interest 48-66, 79, 95, 120, 129
  return on investment (ROI) 78-83
  unemployment 46, 143
rule of three (simple direct, inverse, composite) 18, 37-39, 71, 130

real rate of exchange 36-37
Ricardo, D. 25-26

sequences 61, 139
simple and compound interest 53-56
Sinclair, J. 28
Smith, A. 24-26, 85
Sraffa, P. 25, 26
standard normal distribution 115
statistics 26-30, 67-68, 88-92
stock exchange 118-119
system
    binary 15
    decimal 15
    hexadecimal 15
    decimal metric 15, 23-24, 27
    vigesimal 15
systems
    of measurement 7, 23-24
    numbering 7, 9-16
underlying asset 120-122

net present value (NPV) 81, 95, 129
variables
    discrete and continuous 42
    statistics 29, 92, 115, 116, 132, 133

Walras, Léon 25-26, 102-103, 141
warrants 79, 118, 120

Risks and Hypotheses

The mathematics of economics

A large part of our day-to-day mathematical tasks are related in one way or another to economic science: we compare prices, we calculate the change when we buy something, we read about inflation or unemployment in the news… In fact, it is highly likely that taking out a bank loan or a mortgage will be the most mathematically-challenging decision of our lives. This volume provides an accessible yet rigorous explanation of the mathematics underlying the economy and the finances of people and nations around the world.